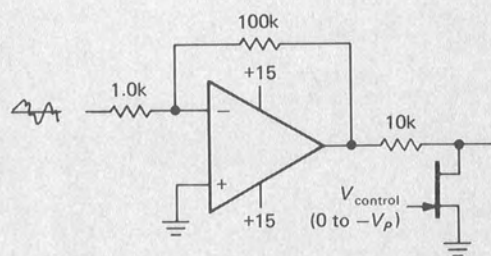
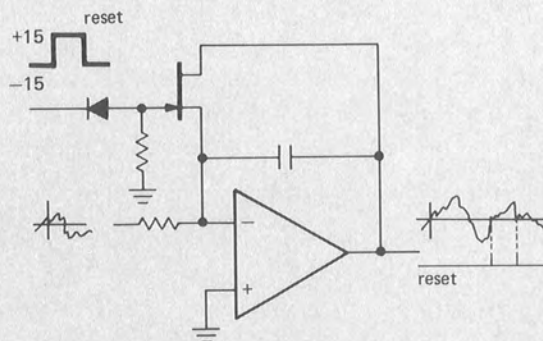


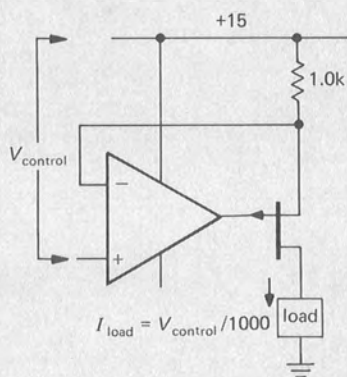
N. op-amps with switches



O. FET gain control



Q. integrator with reset



P. current source

ACTIVE FILTERS AND OSCILLATORS

CHAPTER 5

With only the techniques of transistors and op-amps it is possible to delve into a number of interesting areas of linear (as contrasted with digital) circuitry. We believe that it is important to spend some time doing this now, in order to strengthen your understanding of some of these difficult concepts (transistor behavior, feedback, op-amp limitations, etc.) before introducing more new devices and techniques and getting into the large area of digital electronics. In this chapter, therefore, we will treat briefly the areas of active filters and oscillators. Additional analog techniques are treated in Chapter 6 (voltage regulators and high-current design), Chapter 7 (precision circuits and low noise), Chapter 13 (radiofrequency techniques), Chapter 14 (low-power design), and Chapter 15 (measurements and signal processing). The first part of this chapter (active filters, Sections 5.01–5.11) describes techniques of a somewhat specialized nature, and it can be passed over in a first reading. However, the latter part of this chapter (oscillators, Sections 5.12–5.19) describes techniques of broad utility and should not be omitted.

ACTIVE FILTERS

In Chapter 1 we began a discussion of filters made from resistors and capacitors. Those simple *RC* filters produced gentle high-pass or low-pass gain characteristics, with a 6dB/octave falloff well beyond the -3dB point. By cascading high-pass and low-pass filters, we showed how to obtain bandpass filters, again with gentle 6dB/octave “skirts.” Such filters are sufficient for many purposes, especially if the signal being rejected by the filter is far removed in frequency from the desired signal passband. Some examples are bypassing of radiofrequency signals in audio circuits, “blocking” capacitors for elimination of dc levels, and separation of modulation from a communications “carrier” (see Chapter 13).

5.01 Frequency response with *RC* filters

Often, however, filters with flatter passbands and steeper skirts are needed. This happens whenever signals must be filtered from other interfering signals nearby in

frequency. The obvious next question is whether or not (by cascading a number of identical low-pass filters, say) we can generate an approximation to the ideal “brick-wall” low-pass frequency response, as in Figure 5.1.

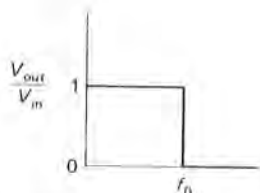
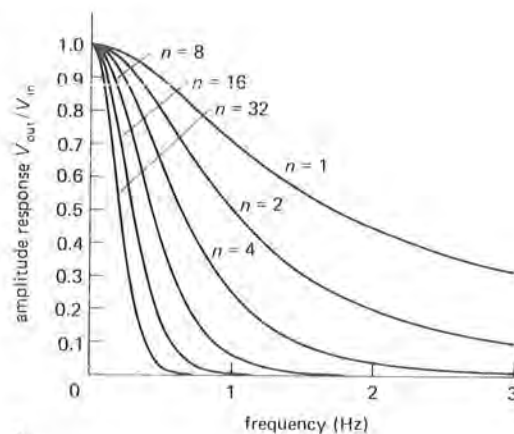


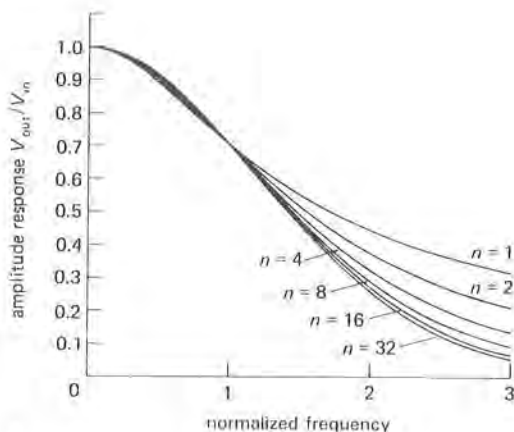
Figure 5.1

We know already that simple cascading won’t work, since each section’s input impedance will load the previous section seriously, degrading the response. But with buffers between each section (or by arranging to have each section of much higher impedance than the one preceding it), it would seem possible. Nonetheless, the answer is no. Cascaded *RC* filters do produce a steep *ultimate* falloff, but the “knee” of the curve of response versus frequency is not sharpened. We might restate this as “many soft knees do not a hard knee make.” To make the point graphically, we have plotted some graphs of gain response (i.e., V_{out}/V_{in}) versus frequency for low-pass filters constructed from 1, 2, 4, 8, 16, and 32 identical *RC* sections, perfectly buffered (Fig. 5.2).

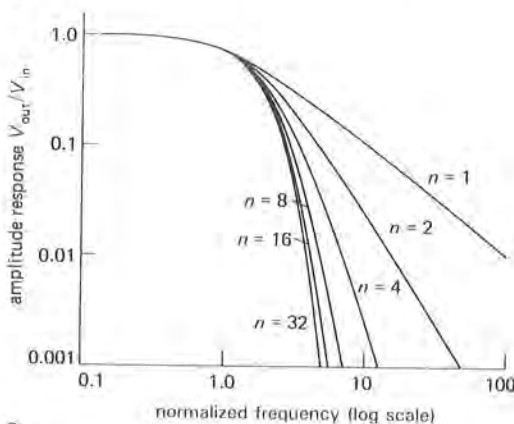
The first graph shows the effect of cascading several *RC* sections, each with its 3dB point at unit frequency. As more sections are added, the overall 3dB point is pushed downward in frequency, as you could easily have predicted. To compare filter characteristics fairly, the rolloff frequencies of the individual sections should be adjusted so that the overall 3dB point is always at the same frequency. The other graphs in Figure 5.2, as well as the next few graphs in this chapter, are all “normalized” in frequency, meaning that the -3 dB point



A



B



C

Figure 5.2. Frequency responses of multisection *RC* filters. Graphs A and B are linear plots, whereas C is logarithmic. The filter responses in B and C have been normalized (or scaled) for 3dB attenuation at unit frequency.

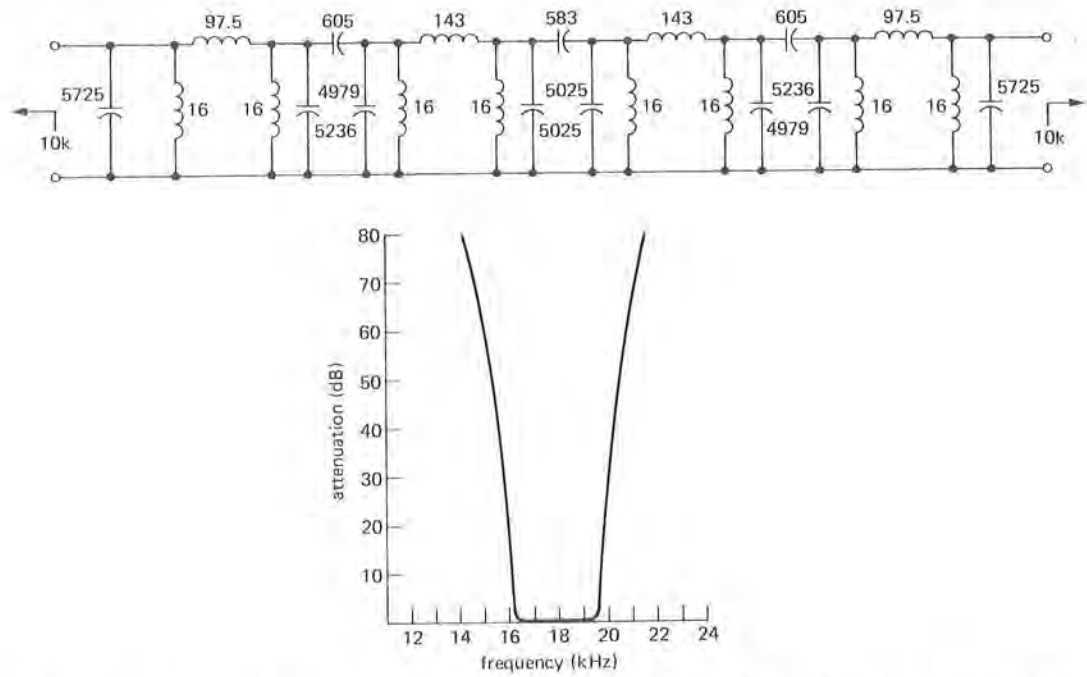


Figure 5.3. An unusually good passive bandpass filter implemented from inductors and capacitors (inductances in mH, capacitances in pF). Bottom: Measured response of the filter circuit. [Based on Figs. 11 and 12 from Orchard, H. J., and Sheahan, D. F., *IEEE Journal of Solid-State Circuits*, Vol. SC-5, No. 3 (1970).]

(or breakpoint, however defined) is at a frequency of 1 radian per second (or at 1 Hz). To determine the response of a filter whose breakpoint is set at some other frequency, simply multiply the values on the frequency axis by the actual breakpoint frequency f_c . In general, we will also stick to the log-log graph of frequency response when talking about filters, because it tells the most about the frequency response. It lets you see the approach to the ultimate rolloff slope, and it permits you to read off accurate values of attenuation. In this case (cascaded RC sections) the normalized graphs in Figures 5.2B and 5.2C demonstrate the soft knee characteristic of passive RC filters.

5.02 Ideal performance with LC filters

As we pointed out in Chapter 1, filters made with inductors and capacitors can

have very sharp responses. The parallel LC resonant circuit is an example. By including inductors in the design, it is possible to create filters with any desired flatness of passband combined with sharpness of transition and steepness of falloff outside the band. Figure 5.3 shows an example of a telephone filter and its characteristics.

Obviously the inclusion of inductors into the design brings about some magic that cannot be performed without them. In the terminology of network analysis, that magic consists in the use of "off-axis poles." Even so, the complexity of the filter increases according to the required flatness of passband and steepness of falloff outside the band, accounting for the large number of components used in the preceding filter. The transient response and phase-shift characteristics are also generally degraded as the amplitude response is improved to

approach the ideal brick-wall characteristic.

The synthesis of filters from passive components (R , L , C) is a highly developed subject, as typified by the authoritative handbook by Zverev (see chapter references at end of book). The only problem is that inductors as circuit elements frequently leave much to be desired. They are often bulky and expensive, and they depart from the ideal by being “lossy,” i.e., by having significant series resistance, as well as other “pathologies” such as nonlinearity, distributed winding capacitance, and susceptibility to magnetic pickup of interference.

What is needed is a way to make inductorless filters with the characteristics of ideal RLC filters.

5.03 Enter active filters: an overview

By using op-amps as part of the filter design, it is possible to synthesize any RLC filter characteristic without using inductors. Such inductorless filters are known as active filters because of the inclusion of an active element (the amplifier).

Active filters can be used to make low-pass, high-pass, bandpass, and band-reject filters, with a choice of filter types according to the important features of the response, e.g., maximal flatness of passband, steepness of skirts, or uniformity of time delay versus frequency (more on this shortly). In addition, “all-pass filters” with flat amplitude response but tailored phase versus frequency can be made (they’re also known as “delay equalizers”), as well as the opposite – a filter with constant phase shift but tailored amplitude response.

□ Negative-impedance converters and gyrators

Two interesting circuit elements that should be mentioned in any overview are the negative-impedance converter (NIC)

and the gyrator. These devices can mimic the properties of inductors, while using only resistors and capacitors in addition to op-amps.

Once you can do that, you can build inductorless filters with the ideal properties of any RLC filter, thus providing at least one way to make active filters.

The NIC converts an impedance to its *negative*, whereas the gyrator converts an impedance to its *inverse*. The following exercises will help you discover for yourself how that works out.

EXERCISE 5.1

Show that the circuit in Figure 5.4 is a negative-impedance converter, in particular that $Z_{in} = -Z$. Hint: Apply some input voltage V , and compute the input current I . Then take the ratio to find $Z_{in} = V/I$.

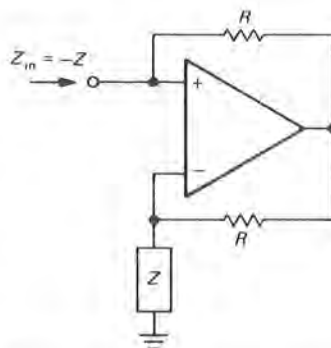


Figure 5.4. Negative-impedance converter.

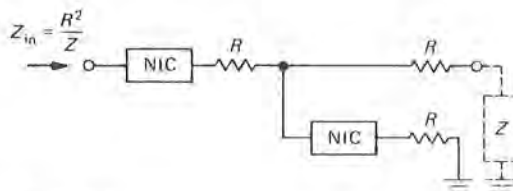


Figure 5.5

EXERCISE 5.2

Show that the circuit in Figure 5.5 is a gyrator, in particular that $Z_{in} = R^2/Z$. Hint: You can analyze it as a set of voltage dividers, beginning at the right.

The NIC therefore converts a capacitor to a “backward” inductor:

$$Z_C = 1/j\omega C \rightarrow Z_{in} = j/\omega C$$

i.e., it is inductive in the sense of generating a current that lags the applied voltage, but its impedance has the wrong frequency dependence (it goes down, instead of up, with increasing frequency). The gyrator, on the other hand, converts a capacitor to a true inductor:

$$Z_C = 1/j\omega C \rightarrow Z_{in} = j\omega CR^2$$

i.e., an inductor with inductance $L = CR^2$.

The existence of the gyrator makes it intuitively reasonable that inductorless filters can be built to mimic any filter using inductors: Simply replace each inductor by a gyrated capacitor. The use of gyrators in just that manner is perfectly OK, and in fact the telephone filter illustrated previously was built that way. In addition to simple gyrator substitution into preexisting *RLC* designs, it is possible to synthesize many other filter configurations. The field of inductorless filter design is extremely active, with new designs appearing in the journals every month.

Sallen-and-Key filter

Figure 5.6 shows an example of a simple and even partly intuitive filter. It is known as a Sallen-and-Key filter, after its inventors. The unity-gain amplifier can be an op-amp connected as a follower, or just an emitter follower. This particular filter is a 2-pole high-pass filter. Note that it would be simply two cascaded *RC* high-pass filters except for the fact that the bottom of the first resistor is bootstrapped by the output. It is easy to see that at very low frequencies it falls off just like a cascaded *RC*, since the output is essentially zero. As the output rises at increasing frequency, however, the bootstrap action tends to reduce

the attenuation, giving a sharper knee. Of course, such hand-waving cannot substitute for honest analysis, which luckily has already been done for a prodigious variety of nice filters. We will come back to active filter circuits in Section 5.06.

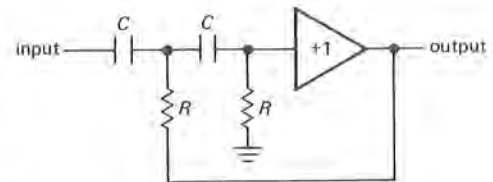


Figure 5.6

5.04 Key filter performance criteria

There are some standard terms that keep appearing when we talk about filters and try to specify their performance. It is worth getting it all straight at the beginning.

Frequency domain

The most obvious characteristic of a filter is its gain versus frequency, typified by the sort of low-pass characteristic shown in Figure 5.7.

The *passband* is the region of frequencies that are relatively unattenuated by the filter. Most often the passband is considered to extend to the -3dB point, but with certain filters (most notably the “equi-ripple” types) the end of the passband may be defined somewhat differently. Within the passband the response may show variations or *ripples*, defining a *ripple band*, as shown. The *cutoff frequency*, f_c , is the end of the passband. The response of the filter then drops off through a *transition region* (also colorfully known as the *skirt* of the filter’s response) to a *stopband*, the region of significant attenuation. The stopband may be defined by some minimum attenuation, e.g., 40dB .

Along with the gain response, the other parameter of importance in the frequency

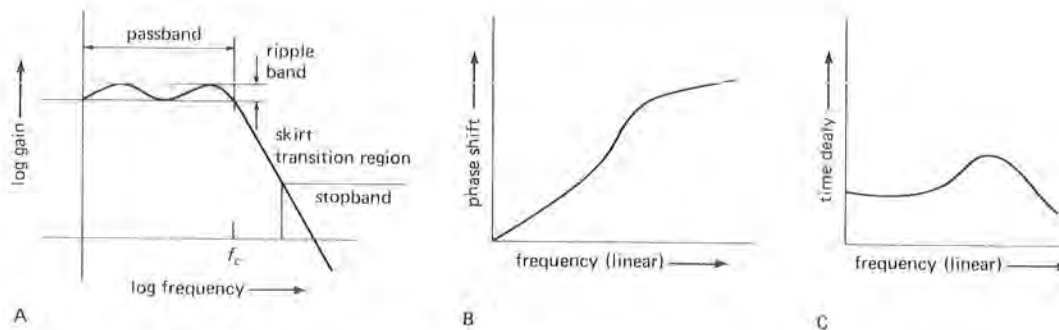


Figure 5.7. Filter characteristics versus frequency.

domain is the *phase shift* of the output signal relative to the input signal. In other words, we are interested in the *complex* response of the filter, which usually goes by the name of $H(s)$, where $s = j\omega$, where H , s , and ω all are complex. Phase is important because a signal entirely within the passband of a filter will emerge with its waveform distorted if the time delay of different frequencies in going through the filter is not constant. Constant time delay corresponds to a phase shift increasing linearly with frequency; hence the term *linear-phase filter* applied to a filter ideal in this respect. Figure 5.8 shows a typical graph of phase shift and amplitude for a low-pass filter that is definitely not a linear-phase filter. Graphs of phase shift versus frequency are best plotted on a linear-frequency axis.

terms for some undesirable properties of filters.

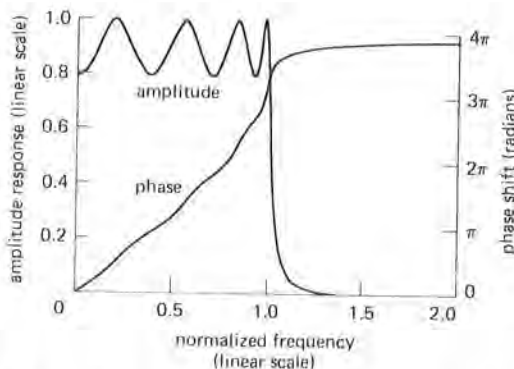


Figure 5.8. Phase and amplitude response for an 8-pole Chebyshev low-pass filter (2dB passband ripple).

5.05 Filter types

Suppose you want a low-pass filter with flat passband and sharp transition to the stopband. The ultimate rate of falloff, well into the stopband, will always be $6n$ dB/octave, where n is the number of "poles." You need one capacitor (or inductor) for each pole, so the required ultimate rate of falloff of filter response determines, roughly, the complexity of the filter.

Now, assume that you have decided to use a 6-pole low-pass filter. You are guaranteed an ultimate rolloff of 36dB/octave at high frequencies. It turns out

Time domain

As with any ac circuit, filters can be described in terms of their *time-domain* properties: rise time, overshoot, ringing, and settling time. This is of particular importance where steps or pulses may be used. Figure 5.9 shows a typical low-pass-filter step response. Here, *rise time* is the time required to reach 90% of the final value, whereas *settling time* is the time required to get within some specified amount of the final value and stay there. *Overshoot* and *ringing* are self-explanatory

sharpness of the knee. A Chebyshev filter is specified in terms of its number of poles and passband ripple. By allowing greater passband ripple, you get a sharper knee. The amplitude is given by

$$\frac{V_{out}}{V_{in}} = \frac{1}{[1 + \epsilon^2 C_n^2(f/f_c)]^{\frac{1}{2}}}$$

where C_n is the Chebyshev polynomial of the first kind of degree n , and ϵ is a constant that sets the passband ripple. Like the Butterworth, the Chebyshev has phase characteristics that are less than ideal.

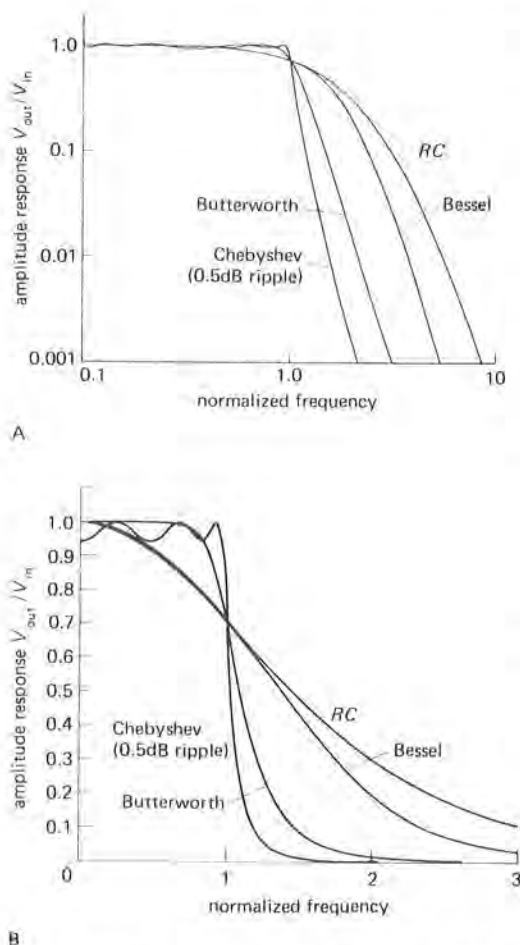


Figure 5.11. Comparison of some common 6-pole low-pass filters. The same filters are plotted on both linear and logarithmic scales.

Figure 5.11 presents graphs comparing the responses of Chebyshev and Butterworth 6-pole low-pass filters. As you can see, they're both tremendous improvements over a 6-pole RC filter.

Actually, the Butterworth, with its maximally flat passband, is not as attractive as it might appear, since you are always accepting some variation in passband response anyway (with the Butterworth it is a gradual rolloff near f_c , whereas with the Chebyshev it is a set of ripples spread throughout the passband). Furthermore, active filters constructed with components of finite tolerance will deviate from the predicted response, which means that a real Butterworth filter will exhibit some passband ripple anyway. The graph in Figure 5.12 illustrates the effects of worst-case variations in resistor and capacitor values on filter response.

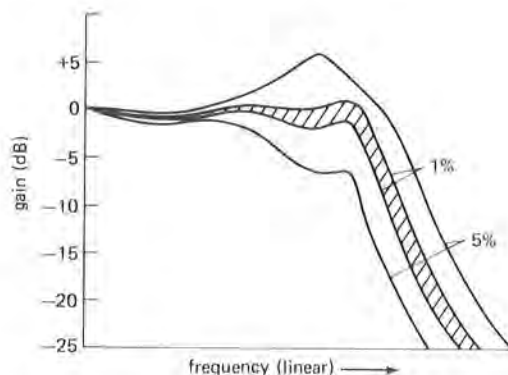


Figure 5.12. The effect of component tolerance on active filter performance.

Viewed in this light, the Chebyshev is a very rational filter design. It is sometimes called an equiripple filter: It manages to improve the situation in the transition region by spreading equal-size ripples throughout the passband, the number of ripples increasing with the order of the filter. Even with rather small ripples (as little as 0.1dB) the Chebyshev filter offers considerably improved sharpness of the knee

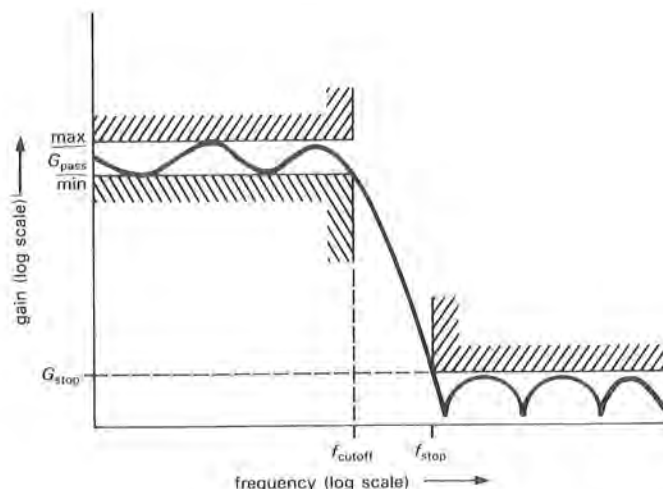


Figure 5.13. Specifying filter frequency response parameters.

as compared with the Butterworth. To make the improvement quantitative, suppose that you need a filter with flatness to 0.1dB within the passband and 20dB attenuation at a frequency 25% beyond the top of the passband. By actual calculation, that will require a 19-pole Butterworth, but only an 8-pole Chebyshev.

The idea of accepting some passband ripple in exchange for improved steepness in the transition region, as in the equiripple Chebyshev filter, is carried to its logical limit in the so-called elliptic (or Cauer) filter by trading ripple in both passband and stopband for an even steeper transition region than that of the Chebyshev filter. With computer-aided design techniques, the design of elliptic filters is as straightforward as for the classic Butterworth and Chebyshev filters.

Figure 5.13 shows how you specify filter frequency response graphically. In this case (a low-pass filter) you indicate the allowable range of filter gain (i.e., the ripple) in the passband, the minimum frequency at which the response leaves the passband, the maximum frequency at which the response enters the stopband, and the minimum attenuation in the stopband.

Bessel filter

As we hinted earlier, the amplitude response of a filter does not tell the whole story. A filter characterized by a flat amplitude response may have large phase shifts. The result is that a signal in the passband will suffer distortion of its waveform. In situations where the shape of the waveform is paramount, a linear-phase filter (or constant-time-delay filter) is desirable. A filter whose phase shift varies linearly with frequency is equivalent to a constant time delay for signals within the passband, i.e., the waveform is not distorted. The Bessel filter (also called the Thomson filter) had maximally flat time delay within its passband, in analogy with the Butterworth, which has maximally flat amplitude response. To see the kind of improvement in time-domain performance you get with the Bessel filter, look at Figure 5.14 for a comparison of time delay versus normalized frequency for 6-pole Bessel and Butterworth low-pass filters. The poor time-delay performance of the Butterworth gives rise to effects such as overshoot when driven with pulse signals. On the other hand, the price you pay for the Bessel's constancy of time delay is an amplitude response

with even less steepness than that of the Butterworth in the transition region between passband and stopband.

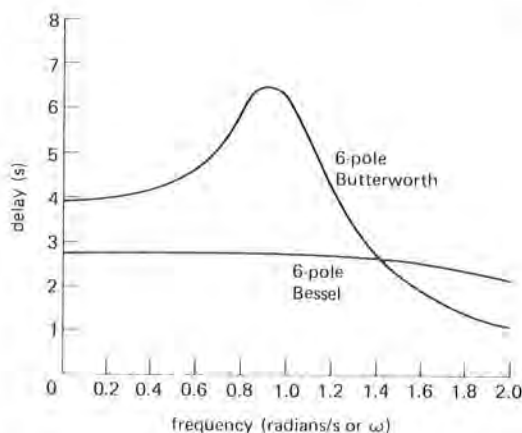


Figure 5.14. Comparison of time delays for 6-pole Bessel and Butterworth low-pass filters. The excellent time-domain performance of the Bessel filter minimizes waveform distortion.

There are numerous filter designs that attempt to improve on the Bessel's good time-domain performance by compromising some of the constancy of time delay for improved rise time and amplitude-versus-frequency characteristics. The Gaussian filter has phase characteristics nearly as good as those of the Bessel, with improved step response. In another class there are interesting filters that allow uniform ripples in the passband time delay (in analogy with the Chebyshev's ripples in its amplitude response) and yield approximately constant time delays even for signals well into the stopband. Another approach to the problem of getting filters with uniform time delays is to use all-pass filters, also known as delay equalizers. These have constant amplitude response with frequency, with a phase shift that can be tailored to individual requirements. Thus, they can be used to improve the time-delay constancy of any filter, including Butterworth and Chebyshev types.

Filter comparison

In spite of the preceding comments about the Bessel filter's transient response, it still has vastly superior properties in the time domain, as compared with the Butterworth and Chebyshev. The Chebyshev, with its highly desirable amplitude-versus-frequency characteristics, actually has the poorest time-domain performance of the three. The Butterworth is in between in both frequency and time-domain properties. Table 5.1 and Figure 5.15 give more information about time-domain performance for these three kinds of filters to complement the frequency-domain graphs presented earlier. They make it clear that the Bessel is a very desirable filter where performance in the time domain is important.

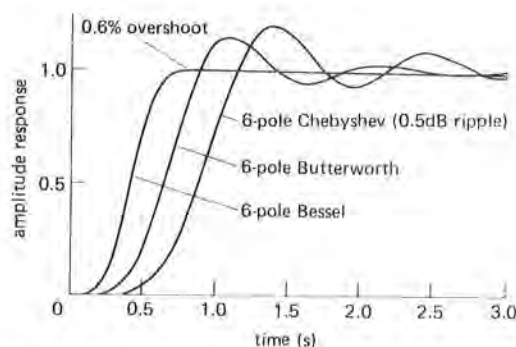


Figure 5.15. Step-response comparison for 6-pole low-pass filters normalized for 3dB attenuation at 1Hz.

ACTIVE FILTER CIRCUITS

A lot of ingenuity has been used in inventing clever active circuits, each of which can be used to generate response functions such as the Butterworth, Chebyshev, etc. You might wonder why the world needs more than one active filter circuit. The reason is that various circuit realizations excel in one or another desirable property, so there is no all-around best circuit.

Some of the features to look for in active filters are (a) small numbers of parts, both

TABLE 5.1. TIME-DOMAIN PERFORMANCE COMPARISON FOR LOW-PASS FILTERS^a

Type	f_{3dB} (Hz)	Poles	Step rise time (0 to 90%) (s)	Over- shoot (%)	Settling time		Stopband attenuation	
					to 1% (s)	to 0.1% (s)	$f = 2f_c$ (dB)	$f = 10f_c$ (dB)
Bessel (-3.0dB at $f_c = 1.0\text{Hz}$)	1.0	2	0.4	0.4	0.6	1.1	10	36
	1.0	4	0.5	0.8	0.7	1.2	13	66
	1.0	6	0.6	0.6	0.7	1.2	14	92
	1.0	8	0.7	0.3	0.8	1.2	14	114
Butterworth (-3.0dB at $f_c = 1.0\text{Hz}$)	1.0	2	0.4	4	0.8	1.7	12	40
	1.0	4	0.6	11	1.0	2.8	24	80
	1.0	6	0.9	14	1.3	3.9	36	120
	1.0	8	1.1	16	1.6	5.1	48	160
Chebyshev 0.5dB ripple (-0.5dB at $f_c = 1.0\text{Hz}$)	1.39	2	0.4	11	1.1	1.6	8	37
	1.09	4	0.7	18	3.0	5.4	31	89
	1.04	6	1.1	21	5.9	10.4	54	141
	1.02	8	1.4	23	8.4	16.4	76	193
Chebyshev 2.0dB ripple (-2.0dB at $f_c = 1.0\text{Hz}$)	1.07	2	0.4	21	1.6	2.7	15	44
	1.02	4	0.7	28	4.8	8.4	37	96
	1.01	6	1.1	32	8.2	16.3	60	148
	1.01	8	1.4	34	11.6	24.8	83	200

^(a) a design procedure for these filters is presented in Section 5.07.

active and passive, (b) ease of adjustability, (c) small spread of parts values, especially the capacitor values, (d) undemanding use of the op-amp, especially requirements on slew rate, bandwidth, and output impedance, (e) the ability to make high- Q filters, and (f) sensitivity of filter characteristics to component values and op-amp gain (in particular, the gain-bandwidth product, f_T). In many ways the last feature is one of the most important. A filter that requires parts of high precision is difficult to adjust, and it will drift as the components age; in addition, there is the nuisance that it requires components of good initial accuracy. The VCVS circuit probably owes most of its popularity to its simplicity and its low parts count, but it suffers from high sensitivity to component variations. By comparison, recent interest in more complicated filter realizations is motivated by the benefits of insensitivity of filter properties to small component variability.

In this section we will present several circuits for low-pass, high-pass, and band-pass active filters. We will begin with the popular VCVS, or controlled-source type, then show the state-variable designs available as integrated circuits from several manufacturers, and finally mention the twin-T sharp rejection filter and some interesting new directions in switched-capacitor realizations.

5.06 VCVS circuits

The voltage-controlled voltage-source (VCVS) filter, also known simply as a controlled-source filter, is a variation of the Sallen-and-Key circuit shown earlier. It replaces the unity-gain follower with a non-inverting amplifier of gain greater than 1. Figure 5.16 shows the circuits for low-pass, high-pass, and bandpass realizations. The resistors at the outputs of the op-amps create a noninverting voltage amplifier

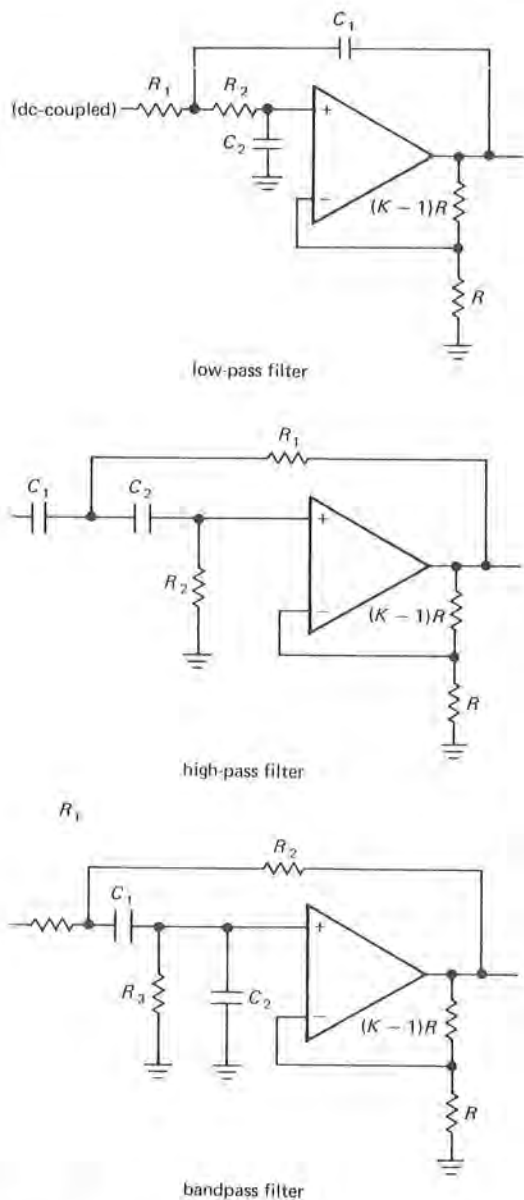


Figure 5.16. VCVS active filter circuits.

of voltage gain K , with the remaining R s and C s contributing the frequency response properties for the filter. These are 2-pole filters, and they can be Butterworth, Bessel, etc., by suitable choice of component values, as we will show later. Any number of VCVS 2-pole sections may be

cascaded to generate higher-order filters. When that is done, the individual filter sections are, in general, not identical. In fact, each section represents a quadratic polynomial factor of the n th-order polynomial describing the overall filter.

There are design equations and tables in most standard filter handbooks for all the standard filter responses, usually including separate tables for each of a number of ripple amplitudes for Chebyshev filters. In the next section we will present an easy-to-use design table for VCVS filters of Butterworth, Bessel, and Chebyshev responses (0.5dB and 2dB passband ripple for Chebyshev filters) for use as low-pass or high-pass filters. Bandpass and band-reject filters can be easily made from combinations of these.

5.07 VCVS filter design using our simplified table

To use Table 5.2, begin by deciding which filter response you need. As we mentioned earlier, the Butterworth may be attractive if maximum flatness of passband is desired, the Chebyshev gives the fastest roll-off from passband to stopband (at the

TABLE 5.2. VCVS LOW-PASS FILTERS

Poles	Butterworth K	Bessel		Chebyshev (0.5dB)		Chebyshev (2.0dB)	
		f_n	K	f_n	K	f_n	K
2	1.586	1.272	1.268	1.231	1.842	0.907	2.114
4	1.152	1.432	1.084	0.597	1.582	0.471	1.924
	2.235	1.606	1.759	1.031	2.660	0.964	2.782
6	1.068	1.607	1.040	0.396	1.537	0.316	1.891
	1.586	1.692	1.364	0.768	2.448	0.730	2.648
	2.483	1.908	2.023	1.011	2.846	0.983	2.904
8	1.038	1.781	1.024	0.297	1.522	0.238	1.879
	1.337	1.835	1.213	0.599	2.379	0.572	2.605
	1.889	1.956	1.593	0.861	2.711	0.842	2.821
	2.610	2.192	2.184	1.006	2.913	0.990	2.946

expense of some ripple in the passband), and the Bessel provides the best phase characteristics, i.e., constant signal delay in the passband, with correspondingly good step response. The frequency responses for all types are shown in the accompanying graphs (Fig. 5.17).

To construct an n -pole filter (n is an even number), you will need to cascade $n/2$ VCVS sections. Only even-order filters are shown, since an odd-order filter requires as many op-amps as the next higher-order filter. Within each section, $R_1 = R_2 = R$, and $C_1 = C_2 = C$. As is usual in op-amp circuits, R will typically be chosen in the range 10k to 100k. (It is best to avoid small resistor values, because the rising open-loop output impedance of the op-amp at high frequencies adds to the resistor values and upsets calculations.) Then all you need to do is set the gain, K , of each stage according to the table entries. For an n -pole filter there are $n/2$ entries, one for each section.

Butterworth low-pass filters

If the filter is a Butterworth, all sections have the same values of R and C , given simply by $RC = 1/2\pi f_c$, where f_c is the desired -3dB frequency of the entire filter. To make a 6-pole low-pass Butterworth filter, for example, you cascade three of the low-pass sections shown previously, with gains of 1.07, 1.59, and 2.48 (preferably in that order, to avoid dynamic range problems), and with identical R s and C s to set the 3dB point. The telescope drive circuit in Section 8.31 shows such an example, with $f_c = 88.4\text{Hz}$ ($R = 180\text{k}$, $C = 0.01\mu\text{F}$).

Bessel and Chebyshev low-pass filters

To make a Bessel or Chebyshev filter with the VCVS, the situation is only slightly more complicated. Again we cascade several 2-pole VCVS filters, with prescribed

gains for each section. Within each section we again use $R_1 = R_2 = R$, and $C_1 = C_2 = C$. However, unlike the situation with the Butterworth, the RC products for the different sections are different and must be scaled by the normalizing factor f_n (given for each section in Table 5.2) according to $RC = 1/2\pi f_n f_c$. Here f_c is again the -3dB point for the Bessel filter, whereas for the Chebyshev filter it defines the end of the passband, i.e., it is the frequency at which the amplitude response falls out of the ripple band on its way into the stopband. For example, the response of a Chebyshev low-pass filter with 0.5dB ripple and $f_c = 100\text{Hz}$ will be flat within $+0\text{dB}$ to -0.5dB from dc to 100Hz , with 0.5dB attenuation at 100Hz and a rapid falloff for frequencies greater than 100Hz . Values are given for Chebyshev filters with 0.5dB and 2.0dB passband ripple; the latter have a somewhat steeper transition into the stopband (Fig. 5.17).

High-pass filters

To make a high-pass filter, use the high-pass configuration shown previously, i.e., with the R s and C s interchanged. For Butterworth filters, everything else remains unchanged (use the same values for R , C , and K). For the Bessel and Chebyshev filters, the K values remain the same, but the normalizing factors f_n must be inverted, i.e., for each section the new f_n equals $1/(f_n)$ listed in Table 5.2).

A bandpass filter can be made by cascading overlapping low-pass and high-pass filters. A band-reject filter can be made by summing the outputs of nonoverlapping low-pass and high-pass filters. However, such cascaded filters won't work well for high- Q filters (extremely sharp bandpass filters) because there is great sensitivity to the component values in the individual (uncoupled) filter sections. In such cases a high- Q single-stage bandpass circuit (e.g., the VCVS bandpass circuit

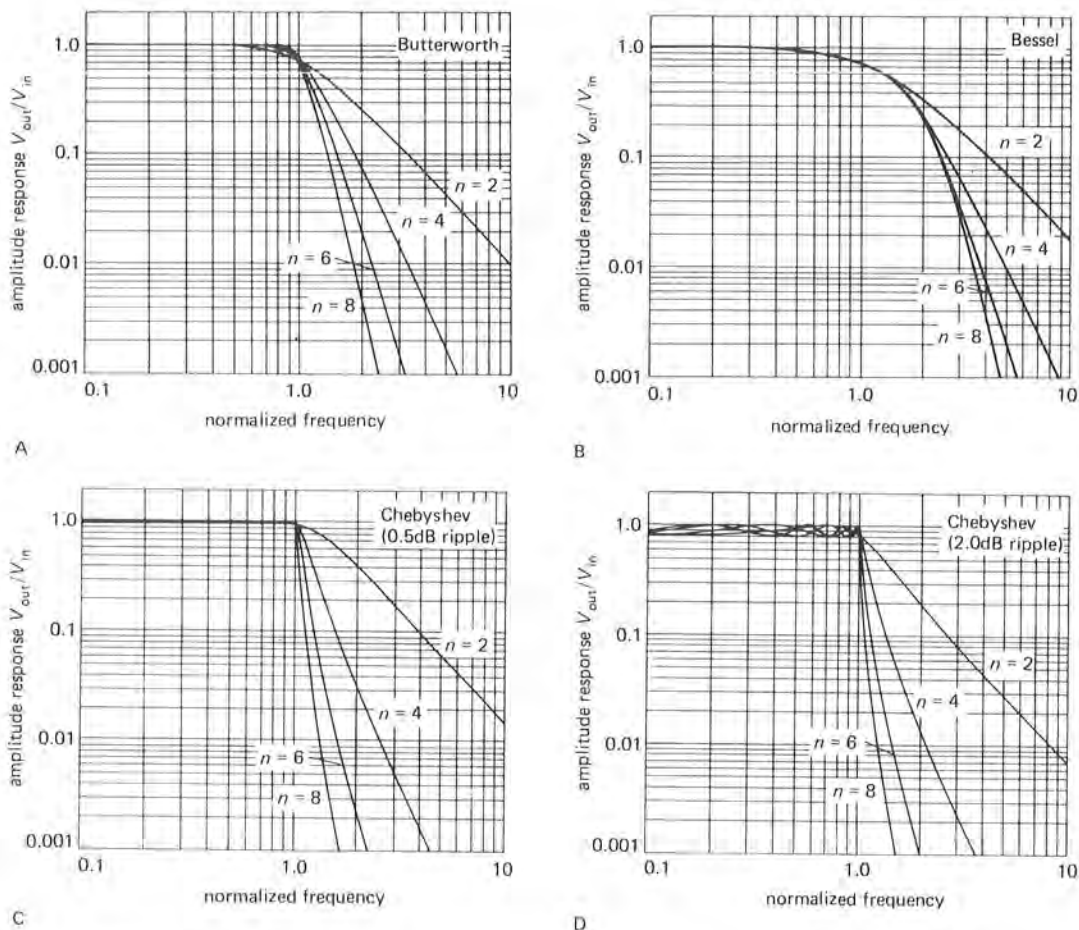


Figure 5.17. Normalized frequency response graphs for the 2-, 4-, 6-, and 8-pole filters in Table 5.2. The Butterworth and Bessel filters are normalized to 3dB attenuation at unit frequency, whereas the Chebyshev filters are normalized to 0.5dB and 2dB attenuations.

illustrated previously, or the state-variable and biquad filters in the next section) should be used instead. Even a single-stage 2-pole filter can produce a response with an extremely sharp peak. Information on such filter design is available in the standard references.

VCVS filters minimize the number of components needed (2 poles/op-amp) and offer the additional advantages of noninverting gain, low output impedance, small spread of component values, easy adjustability of gain, and the ability to operate at high gain or high Q . They suffer from high

sensitivity to component values and amplifier gain, and they don't lend themselves well to applications where a tunable filter of stable characteristics is needed.

EXERCISE 5.3

Design a 6-pole Chebyshev low-pass VCVS filter with a 0.5dB passband ripple and 100Hz cutoff frequency f_c . What is the attenuation at $1.5f_c$?

5.08 State-variable filters

The 2-pole filter shown in Figure 5.18 is far more complex than the VCVS circuits,

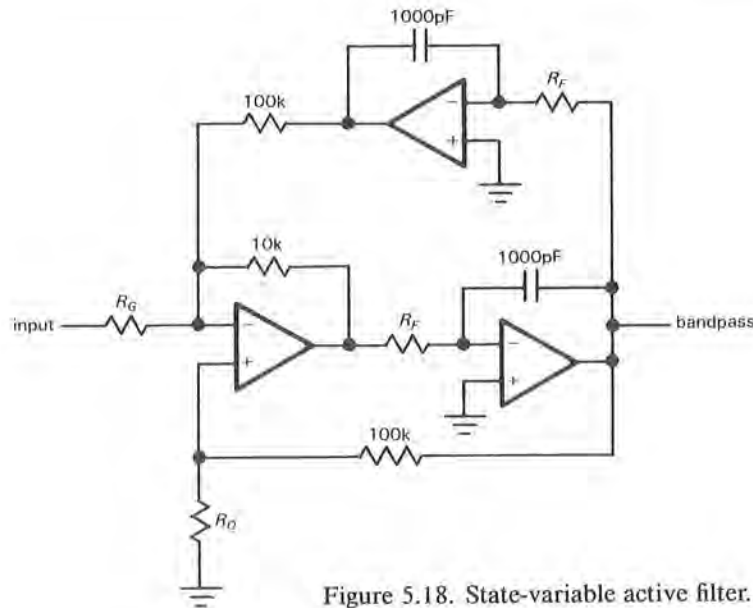


Figure 5.18. State-variable active filter.

but it is popular because of its improved stability and ease of adjustment. It is called a state-variable filter and is available as an IC from National (the AF100 and AF150), Burr-Brown (the UAF series), and others. Because it is a manufactured module, all components except R_G , R_Q , and the two R_F s are built in. Among its nice properties is the availability of high-pass, low-pass, and bandpass outputs from the same circuit; also, its frequency can be tuned while maintaining constant Q (or, alternatively, constant bandwidth) in the bandpass characteristic. As with the VCVS realizations, multiple stages can be cascaded to generate higher-order filters.

Extensive design formulas and tables are provided by the manufacturers for the use of these convenient ICs. They show how to choose the external resistor values to make Butterworth, Bessel, and Chebyshev filters for a wide range of filter orders, for low-pass, high-pass, bandpass, and band-reject responses. Among the nice features of these hybrid ICs is integration of the

capacitors into the module, so that only external resistors need be added.

Bandpass filters

The state-variable circuit, in spite of its large number of components, is a good choice for sharp (high- Q) bandpass filters. It has low component sensitivities, does not make great demands on op-amp bandwidth, and is easy to tune. For example, in the circuit of Figure 5.18, used as a bandpass filter, the two resistors R_F set the center frequency, while R_Q and R_G together determine the Q and band-center gain:

$$R_F = 5.03 \times 10^7 / f_0 \text{ ohms}$$

$$R_Q = 10^5 / (3.48Q + G - 1) \text{ ohms}$$

$$R_G = 3.16 \times 10^4 Q / G \text{ ohms}$$

So you could make a tunable-frequency, constant- Q filter by using a 2-section variable resistor (pot) for R_F . Alternatively, you could make R_Q adjustable, producing a fixed-frequency, variable- Q (and, unfortunately, variable-gain) filter.

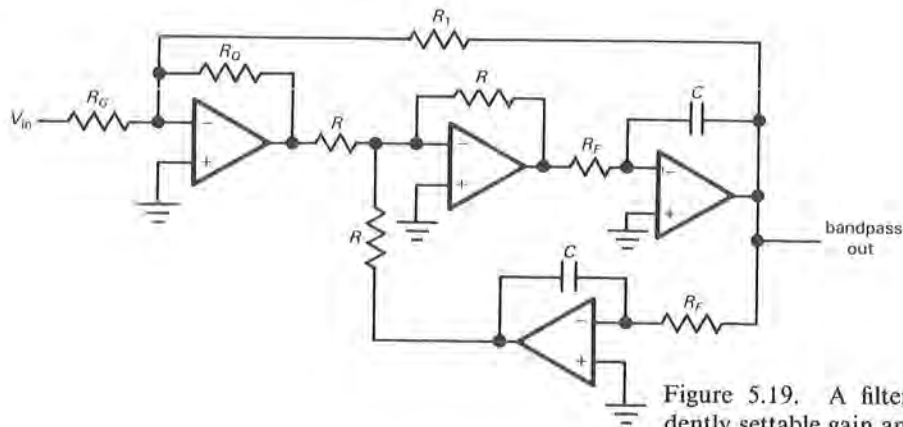


Figure 5.19. A filter with independently settable gain and Q .

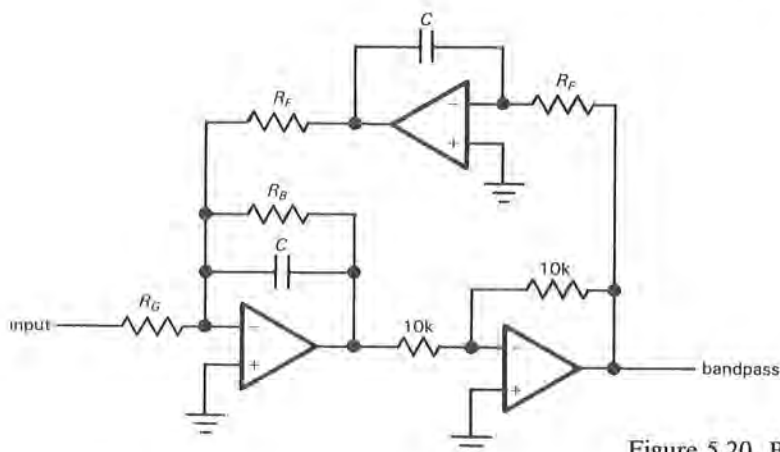


Figure 5.20. Biquad active filter.

EXERCISE 5.4

Calculate resistor values in Figure 5.18 to make a bandpass filter with $f_0 = 1\text{kHz}$, $Q = 50$, and $G = 10$.

Figure 5.19 shows a useful variant of the state-variable bandpass filter. The bad news is that it uses four op-amps; the good news is that you can adjust the bandwidth (i.e., Q) without affecting the midband gain. In fact, both Q and gain are set with a single resistor each. Q , gain, and center frequency are completely independent and are given by these simple equations:

$$\begin{aligned} f_0 &= 1/2\pi R_F C \\ Q &= R_1/R_Q \\ G &= R_1/R_G \\ R &\approx 10\text{k (noncritical, matched)} \end{aligned}$$

Biquad filter. A close relative of the state variable filter is the so-called biquad filter, shown in Figure 5.20. This circuit also uses three op-amps and can be constructed from the state-variable ICs mentioned earlier. It has the interesting property that you can tune its frequency (via R_F) while maintaining constant *bandwidth* (rather than constant Q). Here are the design equations:

$$\begin{aligned} f_0 &= 1/2\pi R_F C \\ \text{BW} &= 1/2\pi R_B C \\ G &= R_B/R_G \end{aligned}$$

The Q is given by f_0/BW and equals R_B/R_F . As the center frequency is varied (via R_F), the Q varies proportionately, keeping the bandwidth Qf_0 constant.

When you design a biquad filter from scratch (rather than with an active filter IC that already contains most of the parts), the general procedure goes something like this:

1. Choose an op-amp whose bandwidth f_T is at least 10 to 20 times Gf_0 .
2. Pick a round-number capacitor value in the vicinity of

$$C = 10/f_0 \mu F$$

3. Use the desired center frequency to calculate the corresponding R_F from the first equation given earlier.
4. Use the desired bandwidth to calculate R_B from the second equation given earlier.
5. Use the desired band-center gain to calculate R_G from the third equation given earlier.

You may have to adjust the capacitor value if the resistor values become awkwardly large or small. For instance, in a high- Q filter you may need to increase C somewhat to keep R_B from becoming too large (or you can use the T-network trick described in Section 4.19). Note that R_F , R_B , and R_G each act as op-amp loads, and should not become less than, say, 5k. When juggling component values, you may find it easier to satisfy requirement 1 by decreasing integrator gain (increase R_F) and simultaneously increasing the inverter-stage gain (increase the 10k feedback resistor).

As an example, suppose we want to make a filter with the same characteristics as in the last exercise. We would begin by provisionally choosing $C = 0.01 \mu F$. Then we find $R_F = 15.9k$ ($f_0 = 1kHz$) and $R_B = 796k$ ($Q = 50$; $BW = 20Hz$). Finally, $R_G = 79.6k$ ($G = 10$).

EXERCISE 5.5

Design a biquad bandpass filter with $f_0 = 60Hz$, $BW = 1Hz$, and $G = 100$.

Higher order bandpass filters

As with our earlier low-pass and high-pass filters, it is possible to build higher order bandpass filters with approximately flat bandpass and steep transition to the stopband.

You do this by cascading several lower-order bandpass filters, the combination tailored to realize the desired filter type (Butterworth, Chebyshev, or whatever). As before, the Butterworth is "maximally flat," whereas the Chebyshev sacrifices passband flatness for steepness of skirts. Both the VCVS and state-variable/biquad bandpass filters just considered are second order (two pole). As you increase the filter sharpness by adding sections, you generally degrade the transient response and phase characteristics. The "bandwidth" of a bandpass filter is defined as the width between $-3dB$ points, except for equiripple filters, for which it is the width between frequencies at which the response falls out of the passband ripple channel.

You can find tables and design procedures for constructing complex filters in standard books on active filters, or in the data sheets for active filter ICs. There are also some very nice filter design programs that run on inexpensive workstations (IBM PC, Macintosh).

□ 5.09 Twin-T notch filters

The passive RC network shown in Figure 5.21 has infinite attenuation at a frequency $f_c = 1/2\pi RC$. Infinite attenuation is

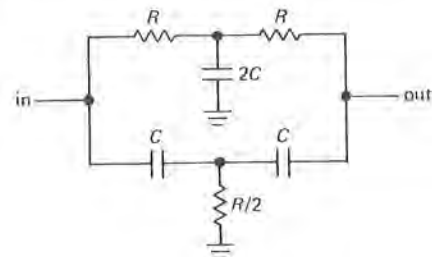


Figure 5.21. Passive twin-T notch filter.

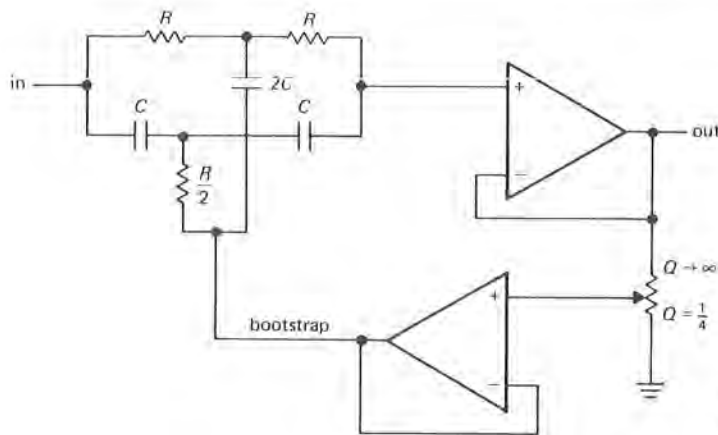


Figure 5.22. Bootstrapped twin-T.

uncharacteristic of RC filters in general; this one works by effectively adding two signals that have been shifted 180° out of phase at the cutoff frequency. It requires good matching of components in order to obtain a good null at f_c . It is called a twin-T, and it can be used to remove an interfering signal, such as 60Hz power-line pickup. The problem is that it has the same “soft” cutoff characteristics as all passive RC networks, except, of course, near f_c , where its response drops like a rock. For example, a twin-T driven by a perfect voltage source is down 10dB at twice (or half) the notch frequency and 3dB at four times (or one-fourth) the notch frequency. One trick to improve its notch characteristic is to “activate” it in the manner of a Sallen-and-Key filter (Fig. 5.22). This technique looks good in principle, but it is generally disappointing in practice, owing to the impossibility of maintaining a good filter null. As the filter notch becomes sharper (more gain in the bootstrap), its null becomes less deep.

Twin-T filters are available as prefab modules, going from 1Hz to 50kHz, with notch depths of about 60dB (with some deterioration at high and low temperatures). They are easy to make from components, but resistors and capacitors of good stability and low temperature coefficient should be used to get a deep and stable notch.

One of the components should be made trimmable.

The twin-T filter works fine as a fixed-frequency notch, but it is a horror to make tunable, since three resistors must be simultaneously adjusted while maintaining constant ratio. However, the remarkably simple RC circuit of Figure 5.23A, which behaves just like the twin-T, can be adjusted over a significant range of frequency (at least two octaves) with a single potentiometer. Like the twin-T (and most active filters) it requires some matching of components; in this case the three capacitors must be identical, and the fixed resistor must be exactly six times the bottom (adjustable) resistor. The notch frequency is then given by

$$f_{\text{notch}} = 1/2\pi C \sqrt{3R_1 R_2}$$

Figure 5.23B shows an implementation that is tunable from 25Hz to 100Hz. The 50k trimmer is adjusted (once) for maximum depth of notch.

As with the passive twin-T, this filter (known as a *bridged differentiator*) has a gently sloping attenuation away from the notch and infinite attenuation (assuming perfect matching of component values) at the notch frequency. It, too, can be “activated,” by bootstrapping the wiper of the pot with a voltage gain somewhat less than unity (as in Fig. 5.22). Increasing

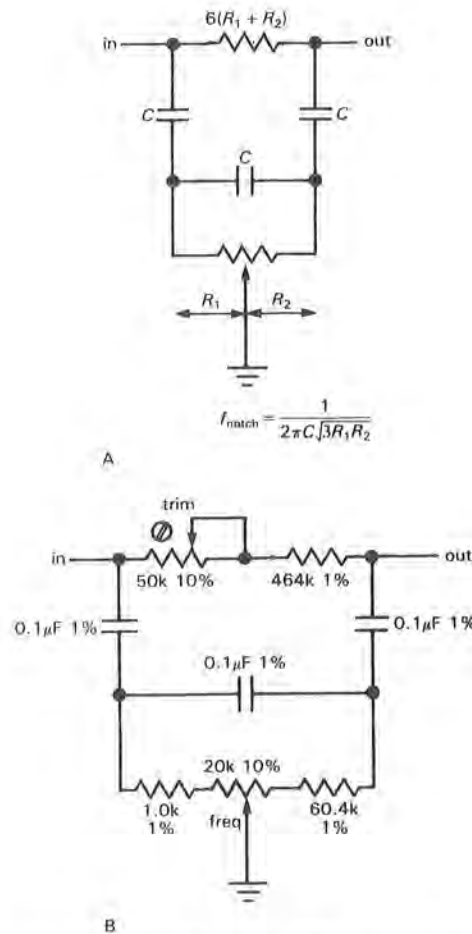


Figure 5.23. Bridged differentiator tunable-notch filter. The implementation in B tunes from 25Hz to 100Hz.

the bootstrap gain toward unity narrows the notch, but also leads to an undesirable response peak on the high frequency side of the notch, along with a reduction in ultimate attenuation.

□ 5.10 Gyrator filter realizations

An interesting type of active filter is made with gyrators; basically they are used to substitute for inductors in traditional filter designs. The gyrator circuit shown in Figure 5.24 is popular. Z_4 will ordinarily be a capacitor, with the other impedances

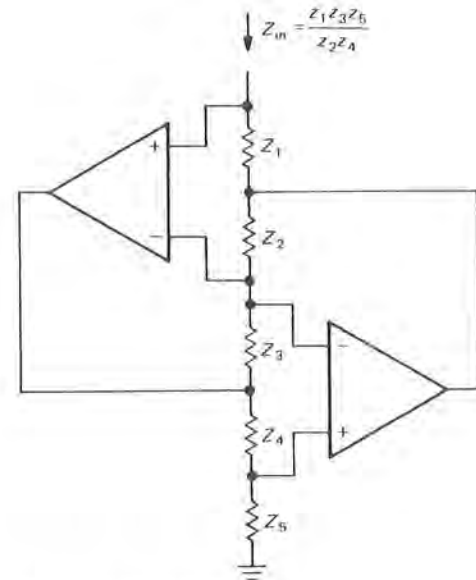
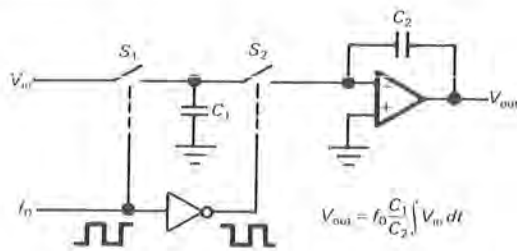


Figure 5.24. Gyrator.

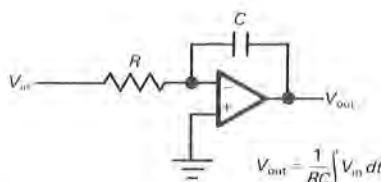
being replaced by resistors, creating an inductor $L = kC$, where $k = R_1 R_3 R_5 / R_2$. It is claimed that these gyrator-substituted filters have the lowest sensitivity to component variations, exactly analogous to their passive *RLC* prototypes.

5.11 Switched capacitor filters

One drawback to these state-variable or biquad filters is the need for accurately matched capacitors. If you build the circuit from op-amps, you've got to get pairs of stable capacitors (not ceramic or electrolytic), perhaps matched as closely as 2% for optimum performance. You also have to make a lot of connections, since the circuits use at least three op-amps and six resistors for each 2-pole section. Alternatively, you can buy a filter IC, letting the manufacturer figure out how to integrate matched 1000pF capacitors into his IC. IC manufacturers have solved those problems, but at a price: The AF100 "Universal Active Filter" IC from National is a hybrid IC and costs about \$10 apiece.



A



B

Figure 5.25. A. Switched-capacitor integrator
B. conventional integrator.

There's another way to implement the integrators that are needed in the state-variable or biquad filter. The basic idea is to use MOS analog switches, clocked from an externally applied square wave at some high frequency (typically 100 times faster than the analog signals of interest), as shown in Figure 5.25. In the figure, the funny triangular object is a digital *inverter*, which turns the square wave upside down so that the two MOS switches are closed on opposite halves of the square wave. The circuit is easy to analyze: When S_1 is closed, C_1 charges to V_{in} , i.e., holding charge $C_1 V_{in}$; on the alternate half of the cycle, C_1 discharges into the virtual ground, transferring its charge to C_2 . The voltage across C_2 therefore changes by an amount $\Delta V = \Delta Q / C_2 = V_{in} C_1 / C_2$. Note that the output voltage *change* during each cycle of the fast square wave is proportional to V_{in} (which we assume changes only a small amount during one cycle of square wave), i.e., the circuit is an integrator! It is easy to show that the integrators obey the equations in the figure.

EXERCISE 5.6

Derive the equations in Figure 5.25

There are two important advantages to using switched capacitors instead of conventional integrators. First, as hinted earlier, it can be less expensive to implement on silicon: The integrator gain depends only on the *ratio* of two capacitors, not on their individual values. In general it is easy to make a matched pair of anything on silicon, but very hard to make a similar component (resistor or capacitor) of precise value and high stability. As a result, monolithic switched-capacitor filter ICs are very inexpensive – National's universal switched-capacitor filter (the MF10) costs \$2 (compared with \$10 for the conventional AF100) and furthermore gives you *two* filters in one package!

The second advantage of switched-capacitor filters is the ability to tune the filter's frequency (e.g., the center frequency of a bandpass filter, or the -3dB point of a low-pass filter) by merely changing the frequency of the square wave ("clock") input. This is because the characteristic frequency of a state-variable or biquad filter is proportional to (and depends only on) the integrator gain.

Switched-capacitor filters are available in both dedicated and "universal" configurations. The former are prewired with on-chip components to form bandpass or low-pass filters, while the latter have various intermediate inputs and outputs brought out so you can connect external components to make anything you want. The price you pay for universality is a larger IC package and the need for external resistors. For example, National's self-contained MF4 Butterworth low-pass filter comes in an 8-pin DIP (\$1.30), while their MF5 universal filter comes in a 14-pin DIP (\$1.45), requiring 2 or 3 external resistors (depending on which filter configuration you choose). Figure 5.26 shows just how easy it is to use the dedicated type.

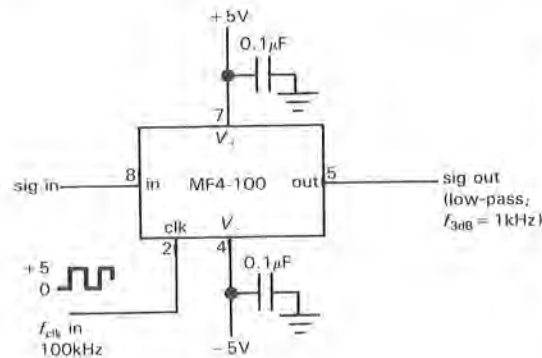


Figure 5.26

Now for the bombshell: Switched-capacitor filters have three annoying characteristics, all related and caused by the presence of the periodic clocking signal. First, there is *clock feedthrough*, the presence of some output signal (typically about 10mV to 25mV) at the clock frequency, independent of the input signal. Usually this doesn't matter, because it is far removed from the signal band of interest. If clock feedthrough is a problem, a simple *RC* filter usually gets rid of it. The second problem is more subtle: If the input signal has any frequency components near the clock frequency, they will be "aliased" down into the passband. To state it precisely, any input signal energy at a frequency that differs from the clock frequency by an amount corresponding to a frequency in the passband will appear (unattenuated!) in the passband. For example, if you use an

MF4 as a 1kHz low-pass filter (i.e., set $f_{\text{clock}} = 100\text{kHz}$), any input signal energy in the range of 99kHz–101kHz will appear in the output band of dc–1kHz. No filter at the output can remove it! You must make sure the input signal doesn't have energy near the clock frequency. If this isn't naturally the case, you can usually use a simple *RC* filter, since the clock frequency is typically quite far removed from the passband. The third undesirable effect in switched-capacitor filters is a general reduction in signal dynamic range (an increase in the "noise floor") due to incomplete cancellation of MOS switch charge injection (see Section 3.12). Typical filter ICs have dynamic ranges of 80dB–90dB.

Like any linear circuit, switched-capacitor filters (and their op-amp analogs) suffer from amplifier errors such as input offset voltage and $1/f$ low-frequency noise. These can be a problem if, for example, you wish to low-pass filter some low-level signal without introducing errors or fluctuations in its average dc value. A nice solution is provided by the clever folks at Linear Technology, who dreamed up the LTC1062 "DC Accurate Low-Pass Filter" (or the MAX280, with improved offset voltage). Figure 5.27 shows how you use it. The basic idea is to put the filter outside the dc path, letting the low-frequency signal components couple passively to the output; the filter grabs onto the signal line only at higher frequencies, where it rolls off the response by shunting the signal to

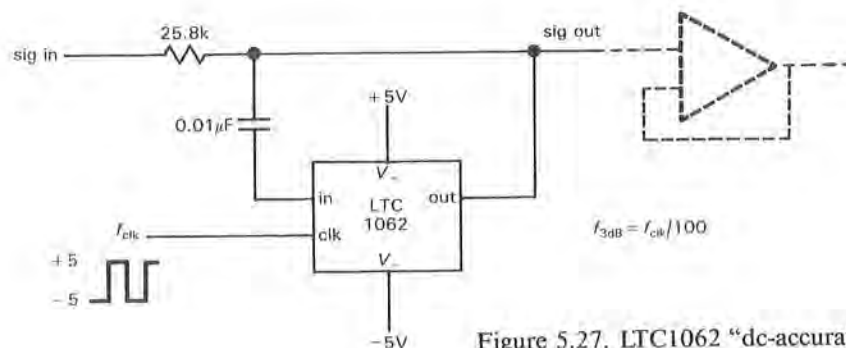


Figure 5.27. LTC1062 "dc-accurate" low-pass filter.

ground. The result is zero dc error, and switched-capacitor-type noise only in the vicinity of the rolloff (Fig. 5.28).

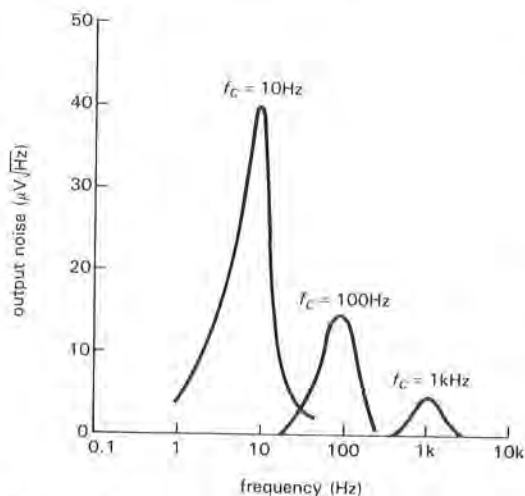


Figure 5.28

Switched-capacitor filter ICs are widely available, from manufacturers such as AMI-Gould, Exar, LTC, National, and EGG-Reticon. Typically you can put the cutoff (or band center) anywhere in the range of dc to a few tens of kilohertz, as set by the clock frequency. The characteristic frequency is a fixed multiple of the clock, usually $50f_{clk}$ or $100f_{clk}$. Most switched-capacitor filter ICs are intended for low-pass, bandpass, or notch (band-stop) use, though a few (e.g., the AMI 3529) are designed as high-pass filters. Note that clock feedthrough and discrete (clock frequency) output waveform quantization effects are particularly bothersome in the latter case, since they're both in-band.

OSCILLATORS

5.12 Introduction to oscillators

Within nearly every electronic instrument it is essential to have an oscillator or waveform generator of some sort. Apart from

the obvious case of signal generators, function generators, and pulse generators themselves, a source of regular oscillations is necessary in any cyclical measuring instrument, in any instrument that initiates measurements or processes, and in any instrument whose function involves periodic states or periodic waveforms. That includes just about everything. For example, oscillators or waveform generators are used in digital multimeters, oscilloscopes, radiofrequency receivers, computers, every computer peripheral (tape, disk, printer, alphanumeric terminal), nearly every digital instrument (counters, timers, calculators, and anything with a "multiplexed display"), and a host of other devices too numerous to mention. A device without an oscillator either doesn't do anything or expects to be driven by something else (which probably contains an oscillator). It is not an exaggeration to say that an oscillator of some sort is as essential an ingredient in electronics as a regulated supply of dc power.

Depending on the application, an oscillator may be used simply as a source of regularly spaced pulses (e.g., a "clock" for a digital system), or demands may be made on its stability and accuracy (e.g., the time base for a frequency counter), its adjustability (e.g., the local oscillator in a transmitter or receiver), or its ability to produce accurate waveforms (e.g., the horizontal-sweep ramp generator in an oscilloscope).

In the following sections we will treat briefly the most popular oscillators, from the simple RC relaxation oscillators to the stable quartz-crystal oscillators. Our aim is not to survey everything in exhaustive detail, but simply to make you acquainted with what is available and what sorts of oscillators are suitable in various situations.

5.13 Relaxation oscillators

A very simple kind of oscillator can be made by charging a capacitor through a

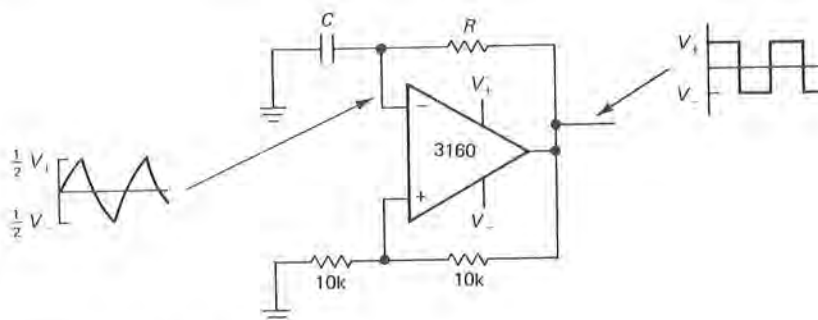


Figure 5.29. Op-amp relaxation oscillator.

resistor (or a current source), then discharging it rapidly when the voltage reaches some threshold, beginning the cycle anew. Alternatively, the external circuit may be arranged to reverse the polarity of the charging current when the threshold is reached, thus generating a triangle wave rather than a sawtooth. Oscillators based on this principle are known as relaxation oscillators. They are inexpensive and simple, and with careful design they can be made quite stable in frequency.

In the past, negative-resistance devices such as unijunction transistors and neon bulbs were used to make relaxation oscillators, but current practice favors op-amps or special timer ICs. Figure 5.29 shows a classic RC relaxation oscillator. The operation is simple: Assume that when power is first applied, the op-amp output goes to positive saturation (it's actually a toss-up which way it will go, but it doesn't matter). The capacitor begins charging up toward V_+ , with time constant RC . When it reaches one-half the supply voltage, the op-amp switches into negative saturation (it's a Schmitt trigger), and the capacitor begins discharging toward V_- with the same time constant. The cycle repeats indefinitely, with period $2.2RC$, independent of supply voltage. A CMOS output-stage op-amp (see Sections 4.11 and 4.22) was chosen because its outputs saturate cleanly at the supply voltages. The bipolar LM10 also swings rail-to-rail and, unlike CMOS op-

amps, allows operation at a full ± 15 volts; however, it has a much lower f_T (0.1MHz).

EXERCISE 5.7

Show that the period is as stated.

By using current sources to charge the capacitor, a good triangle wave can be generated. A clever circuit using that principle was shown in Section 4.29.

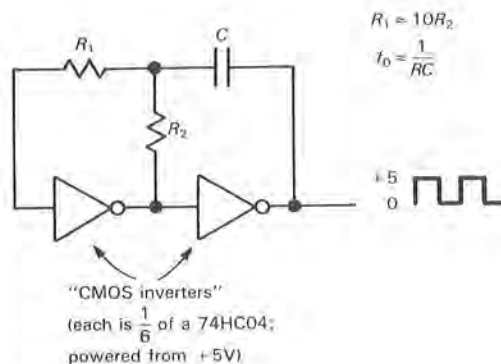


Figure 5.30

Sometimes you need an oscillator with very low noise content (also called "low sideband noise"). The simple circuit of Figure 5.30 is good in this respect. It uses a pair of CMOS inverters (a form of digital logic we'll use extensively in Chapters 8–11) connected together to form an RC relaxation oscillator with square wave output. Actual measurements



Figure 5.31. Low-noise oscillator.

The circuit of Figure 5.31 has even lower noise and furthermore lets you modulate the output frequency via an external current applied to the base of Q_1 . In this circuit Q_1 operates as an integrator, generating an asymmetrical triangle waveform at its collector. The inverters operate as a noninverting comparator, alternating the polarity of the base drive each half cycle. This circuit has close-in noise density of $-90\text{dBc}/\sqrt{\text{Hz}}$ measured 100Hz from the 150kHz carrier, and $-100\text{dBc}/\sqrt{\text{Hz}}$ measured at an offset of 300Hz. Although these circuits excel in low sideband noise, the oscillation frequency has more supply-voltage sensitivity than other oscillators discussed in this chapter.

5.14 The classic timer chip: the 555

The easiest way to understand the workings of the 555 is to look at an example (Fig. 5.33). When power is applied, the capacitor is discharged; so the 555 is triggered, causing the output to go HIGH, the discharge transistor Q_1 to turn

single pulses of arbitrary width, as well as a bunch of other things. It is really a small kit, containing comparators, gates, and flip-flops. It has become a game in the electronics industry to try to think of new uses for the 555. Suffice it to say that many succeed at this new form of entertainment.

A caution about the 555: The 555, along with some other timer chips, generates a big ($\approx 150\text{mA}$) supply-current glitch during each output transition. Be sure to use a hefty bypass capacitor near the chip. Even so, the 555 may have a tendency to generate double output transitions.

CMOS 555s

Some of the less desirable properties of the 555 (high supply current, high trigger current, double output transitions, and inability to run with very low supply voltage) have been remedied in a collection of CMOS successors. You can recognize these by the telltale “555” somewhere in the part number. Table 5.3 lists most of these that we could find, along with their important specifications. Note particularly the ability to operate at very low supply voltage (down to 1V!) and the generally low supply current. These chips also can run at higher frequency than the original 555. The CMOS output stages give rail-to-rail swing, at least at low load currents (but note that these chips don’t have the output-current muscle of the standard 555). All chips listed are CMOS except for the original 555 and the XR-L555. The latter is intended as a bipolar low-power 555 and reveals its pedigree by the hefty output sourcing capability and good tempo.

The 555 oscillator of Figure 5.33 generates a rectangular-wave output whose duty cycle (fraction of time the output is HIGH) is always greater than 50%. That is because the timing capacitor is charged through the series pair $R_A + R_B$, but

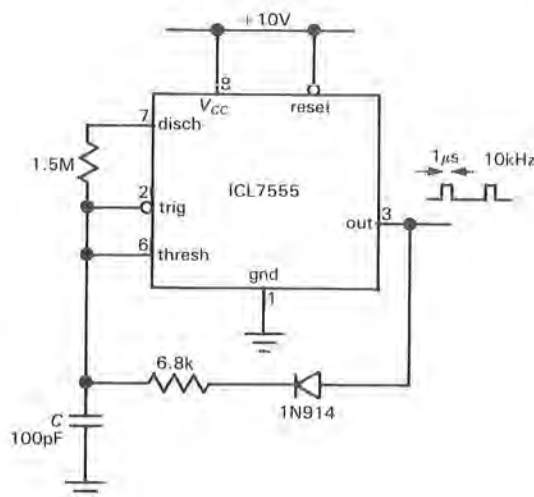


Figure 5.34. Low-duty-cycle oscillator.

discharged (more rapidly) through R_B alone. Figure 5.34 shows how to trick the 555 into giving you low duty-cycle positive pulses. The diode/resistor combination charges timing capacitor rapidly via the output, with slower discharge via the internal discharge transistor. You can only play this trick with a CMOS 555, because you need the full positive output swing.

By using a current source to charge the timing capacitor, you can make a ramp (or “sawtooth-wave”) generator. Figure 5.35 shows how, using a simple *pnp* current source. The ramp charges to $\frac{2}{3}V_{CC}$, then discharges rapidly (through the 555’s *nnp* discharge transistor, pin 7) to $\frac{1}{3}V_{CC}$, beginning the ramp cycle anew. Note that the ramp waveform appears on the capacitor terminal and must be buffered with an op-amp since it is at high impedance. In this circuit you could simplify things somewhat by using a JFET “current-regulator diode” (Section 3.06) in place of the *pnp* current source; however, the performance (ramp linearity) would be slightly degraded, because a JFET operating at I_{DSS} is not as good a current source as the bipolar transistor circuit.

Figure 5.36 shows a simple way to

TABLE 5.3. 555-TYPE OSCILLATORS

Type	Mfg ^a	Qty per package			Supply voltage		Supply curr per osc (V _S = 5V)		Trig, thresh current		Max freq (V _S = 5V)		Tempco type (ppm/°C)	V _{sat} , typ			Rail to rail? ^b	I _{out} , max (V _S =5V, V _O =2.5V)	
		1	2	4	min (V)	max (V)	typ (μA)	max (μA)	typ (nA)	max (nA)	min (MHz)	typ (MHz)		V _{OH} @ I _{src} (V) (mA)	V _{OL} @ I _{sink} (V) (mA)	source (mA)		sink (mA)	
555	SN+	*	*	—	4.5	18	3000	5000	100	500	—	0.5	30	1.4	2	0.1	10	200	200
ICL7555	IL	*	*	—	2	18	60	300	—	10	—	1	150	1	2	0.5	10	4	25
TLC551	TI	*	*	—	1	18	170	—	0.01	—	—	2.1 ^c	—	1	2	0.2	10	—	—
TLC555	TI	*	*	—	2	18	170	—	0.01	—	—	2.1	—	1	2	0.2	10	—	—
LMC555	NS	*	—	—	1.5	15	100	250	0.01	—	—	3	75	0.3	2	0.3	10	—	—
ALD555-1	AL	*	—	—	1	12	100	180	0.001	0.2	1.4	2	300	0.4	2	0.2	10	3	100
ALD1504	AL	*	*	—	1	12	50	90	0.01	0.4	1.5	2.5	300	0.4	2	0.2	10	10	100
ALD4503	AL	—	—	*	1	12	35	70	0.01	0.4	—	2	300	0.4	2	0.2	10	3	100
XR-L555M	XR	*	*	—	2.7	15	150	300	500	—	—	—	30	1.7	10	0.3	2	100	—

(a) see footnotes to Table 4.1. (b) signifies that the output stage can swing to both rails. (c) at V_S=1.2V.

(a) see footnotes to Table 4.1. (b) signifies that the output stage can swing to both rails. (c) at V_S = 1.2V.

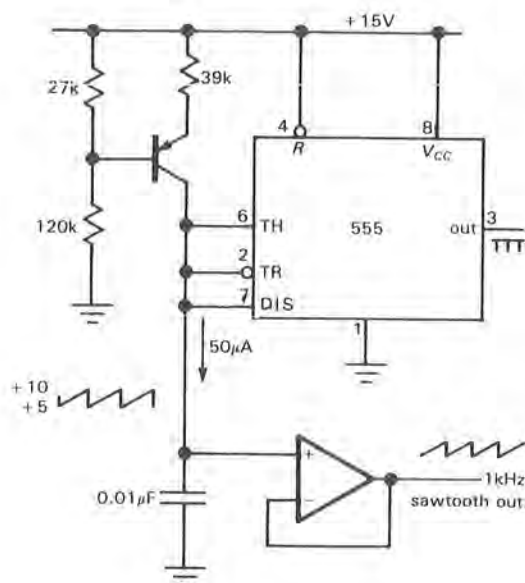


Figure 5.35. Sawtooth oscillator.

behaves like a normal diode in the reverse direction, owing to gate-drain conduction). The rail-to-rail output swing thus generates a constant current, of alternating polarity, producing a triangle waveform (going between the usual $\frac{1}{3}V_{CC}$ and $\frac{2}{3}V_{CC}$) at the capacitor. As before, you have to buffer the high-impedance waveform with an op-amp. Note that you must use a CMOS 555, particularly when operating the circuit from +5 volts, since the circuit depends on a full rail-to-rail output swing. For example, the HIGH output of a bipolar 555 is typically 2 diode drops below the positive rail (*npn* Darlington follower), or +3.8 volts with a 5 volt supply; this leaves only 0.5 volt across the series pair of current regulators at the top of the waveform, obviously insufficient to turn on the current regulator (approximately 1V) and the series JFET diode (0.6V).

generate a *triangle* wave with a CMOS 555. Here we wired a pair of JFET current regulators in series to generate a bidirectional current regulator (each current regulator

EXERCISE 5.9

Demonstrate that you understand the circuits of Figures 5.35 and 5.36 by calculating the frequency of oscillation in each case.

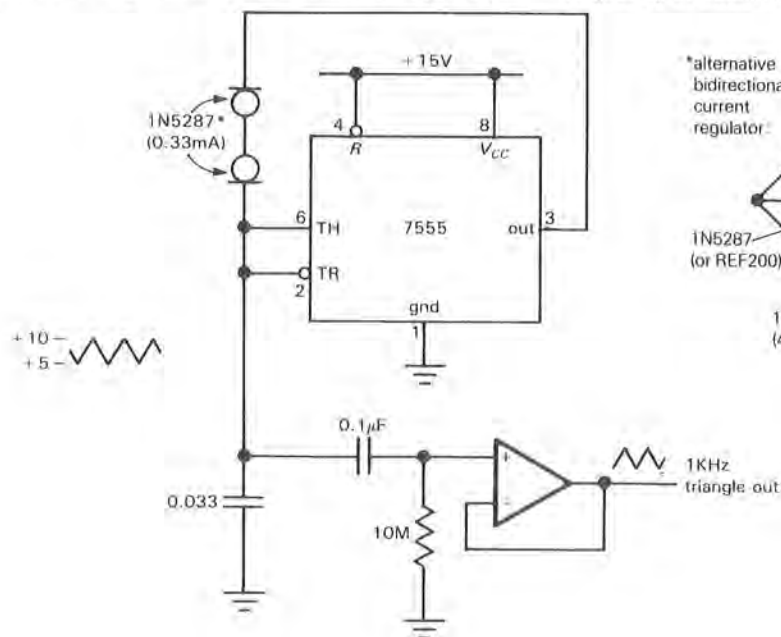
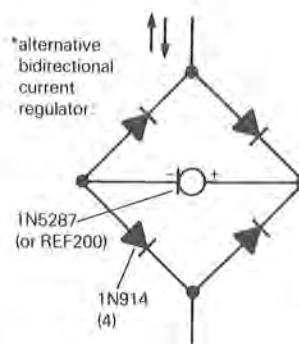


Figure 5.36. Triangle generator.



There are some other interesting timer chips available. The 322 timer from National includes its own internal precision voltage reference for determining the threshold. That makes it an excellent choice for generating a frequency proportional to an externally supplied current, as, for example, from a photodiode. Another class of timers uses a relaxation oscillator followed by a digital counter, in order to generate long delay times without resorting to large resistor and capacitor values. Examples of this are the 74HC4060, the Exar 2243, and the Intersil ICM7242 (also made by Maxim). The latter is CMOS, runs on a fraction of a milliamp, and generates an output pulse every 128 oscillator cycles. These timers (and their near relatives) are great for generating delays from a few seconds to a few minutes.

5.15 Voltage-controlled oscillators

Other IC oscillators are available as voltage-controlled oscillators (VCO's), with the output rate variable over some range according to an input control voltage. Some of these have frequency ranges exceeding 1000:1. Examples are the original NE566 and later designs like the LM331, 8038, 2206, and 74LS624-9 series.

The 74LS624 series, for example, generates digital-logic-level outputs up to 20 MHz and uses external RC s to set the nominal frequency. Faster VCOs like the 1648 can produce outputs to 200MHz, and in Chapter 13 we'll see how to make VCOs that operate in the *gigahertz* range. The LM331 is actually an example of a voltage-to-frequency (V/F) converter, designed for good linearity (see Sections 9.20 and 9.27). Where linearity is important, recent V/F converters like the AD650 really do the job, with linearity of 0.005%. Most VCOs use internal current sources to generate triangle-wave outputs, and the 8038 and 2206 even include a set of "soft" clamps to convert the triangle wave to a

not-too-great sine wave. VCO chips sometimes have an awkward reference for the control voltage (e.g., the positive supply) and complicated symmetrizing schemes for sine-wave output. It is our opinion that the ideal VCO has yet to be developed. Many of these chips can be used with an external quartz crystal, as we will discuss shortly, for much higher accuracy and stability; in such cases the crystal simply replaces the capacitor. Figure 5.37 shows a VCO circuit with an output frequency range of 10Hz to 10kHz built with the LM331.

When shopping for VCO chips, don't overlook the ICs known as *phase-locked loops* (PLL), which contain both a VCO and a phase detector. An example is the popular CMOS 4046 (and its faster cousin, the 74HC4046). We will discuss PLLs in Sections 9.27-9.31. Table 5.4 lists most of the available VCOs.

5.16 Quadrature oscillators

There are times when you need an oscillator that generates a simultaneous *pair* of equal-amplitude sine waves, *90° out of phase*. You can think of the pair as sine and cosine. This is referred to as a *quadrature pair* (the signals are "in quadrature"). One important application is in radio communications circuits (quadrature mixers, single-sideband generation). Furthermore, as we'll explain below, a quadrature pair is all you need to generate any arbitrary phase.

The first idea you might invent is to apply a sine-wave signal to an integrator (or differentiator), thus generating a 90°-shifted cosine wave. The phase shift is right, but the amplitude is wrong (figure out why). Here are some methods that do work:

□ Switched-capacitor resonator

Figure 5.38 shows how to use an MF5 switched-capacitor filter IC as a self-excited

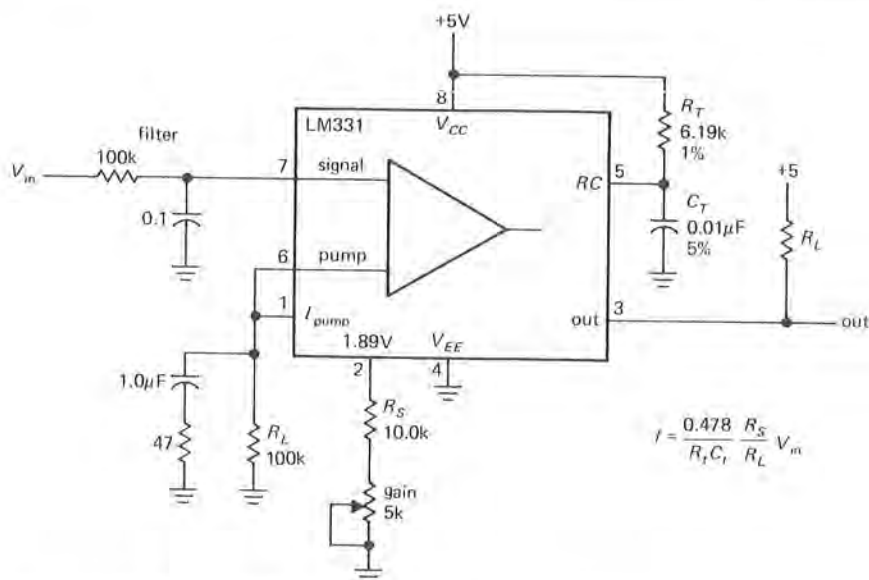


Figure 5.37. Typical V/F converter IC (0 to 10kHz VCO).

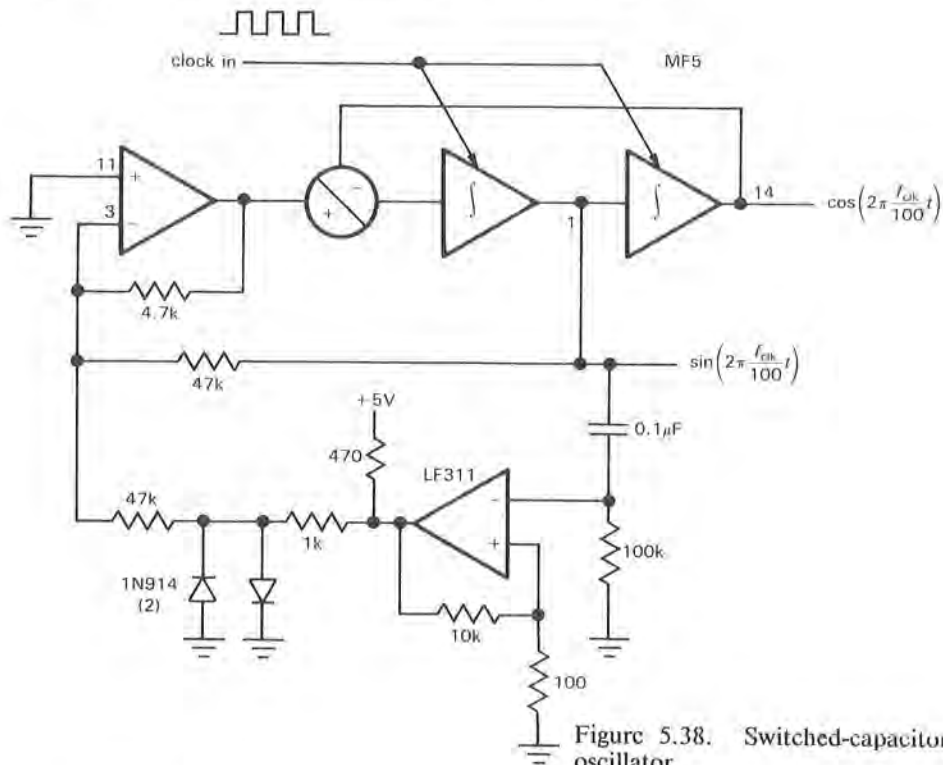


Figure 5.38. Switched-capacitor quadrature oscillator.

bandpass filter to generate a quadrature sine-wave pair. The easiest way to under-

stand it is to assume there is already a sine-wave output present; the comparator

TABLE 5.4. SELECTED VCOs

Type	Mfg ^a	Family ^b	Max freq (MHz)	Outputs ^c	Supply voltage		Linearity (at 10kHz)	Comments
					min (V)	max (V)		
VFC32	BB+	L	0.5	OC	±9	±18	0.01%	indus. st'd; good linearity
VFC62C	BB	L	1	OC	±13	±20	0.002%	excellent linearity
VFC110B	BB	L	4	OC	±8	±18	0.005%	fast, exc lin, int V_{ref}
74S124	TI	T	60	SQ	4.75	5.25		
74LS624-9	TI	T	20	SQ	4.75	5.25		
74LS724	TI	T	16	SQ	4.75	5.25		mini-DIP
215	XR	L	35	SQ	5	26		PLL
LM331	NS	L	0.1	OC	4	40		inexpensive, good linearity
AD537	AD	L	0.1	OC	4.5	36	0.07%	
566	SN	L	1	SQ,T	10	24		
AD650	AD	L	1	OC	±9	±18	0.005%	excellent linearity
AD654	AD	L	0.5	OC	4.5	36	0.1% ^d	inexpensive
1648	MO	E	200	P	-5.2			
1658	MO	E	130	P	-5.2			
XR2206	XR	L	0.5	SQ,T,SW	10	26	2%	0.5% sine dist (trimmed)
XR2207	XR	L	0.5	SQ,T	8	26	1%	
XR2209	XR	L	1	SQ,T	±4	±13	1%	
XR2212	XR	L	0.3	SQ	4.5	20		PLL
XR2213	XR	L	0.3	SQ	4.5	15		PLL
4024	MO	T	25	SQ	4.75	5.25		
4046	RC+	C	1	SQ	3	15		CMOS PLL
HC4046	RC+	C	15 [†]	SQ	3	6		fast 4046
4151	RA	L	0.1	OC	8	22	0.013%	
4152	RA	L	0.1	OC	7	18	0.007%	
4153A	RA	L	0.5	OC	±12	±18	0.002%	excellent lin, easy to use
8038	IL	L	0.1	SQ,T,SW	10	30	0.2%	Exar 8038 to 1MHz
TSC9401	TP	L	0.1	OC	±4	±7.5	0.01%	V/F, linear, stable

(a) see footnotes to Table 4.1. (b) families: C – CMOS; E – ECL; L – linear; T – TTL. (c) outputs: OC – open collector, pulses; P – pulses; SQ – square waves; SW – sine waves; T – triangle. (d) at 250kHz.

converts this to a small-amplitude (1 diode drop) square wave, which is fed back as the filter's input. The filter has a narrow bandpass ($Q = 10$), so it converts the input square wave to a sine-wave output, sustaining the oscillation. A square-wave clock input (CLK) determines the bandpass center frequency, hence the frequency of oscillation, in this case $f_{clk}/100$. The circuit is usable over a frequency range of a few hertz to about 10kHz and generates a quadrature pair of sine waves of equal amplitude. Note that this circuit will actually have a "staircase" approximation to the desired sine-wave output, owing to

the quantized output steps of the switched filter.

□ Analog trigonometric-function generator

Analog Devices makes an interesting non-linear "function IC" that converts an input voltage to an output voltage proportional to $\sin(AV_{in})$, where the gain A is fixed at 50° per volt. In fact, this chip, the AD639, can actually do a lot more: It has four inputs, called X_1 , X_2 , Y_1 , and Y_2 , and generates as output the voltage $V_{out} = \sin(X_1 - X_2)/\sin(Y_1 - Y_2)$. Thus, for example, by setting $X_1 = Y_1 = 90^\circ$

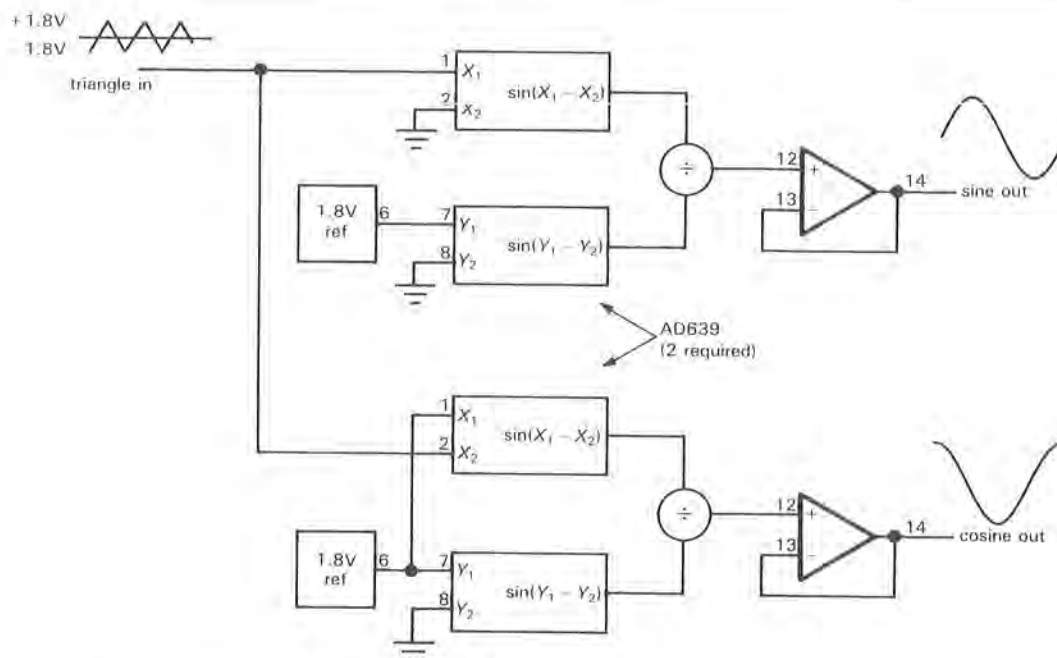


Figure 5.39. Trigonometric-function oscillator.

(i.e., +1.8 volt), $Y_2 = 0$ (ground), and applying an input voltage to X_2 , we generate $\cos(X_2)$.

EXERCISE 5.10

Prove the last statement.

The AD639 even gives you a precise +1.8 volt output, to make life easy. Thus, a pair of AD639s, driven by a 1.8 volt amplitude triangle wave, generates a quadrature sine-wave pair, as shown in Figure 5.39. The AD639 operates from dc to about 1MHz.

□ Lookup table

This is a digital technique, which you will fully understand only after you've read Chapter 9. The idea is to program a digital memory with the numerical values of sine and cosine for a large set of equally spaced angle arguments (say for every 1°). You then make sine waves by rapidly generating the sequential addresses, reading the

memory values for each address (i.e., each sequential angle), and applying the digital values to a pair of digital-to-analog (D/A) converters.

This method has some drawbacks. As with the switched-capacitor resonator, the output is actually a staircase wave, since it is constructed from a set of discrete voltages, one for each table entry. You can, of course, use a low-pass filter to smooth the output; but having done so, you cannot span a wide range of frequencies, since the low-pass filter must be chosen to pass the sine wave itself while blocking the (higher) angle step frequency (the same problem applies to the switched-capacitor resonator). Decreasing the angular step size helps, but reduces the maximum output frequency. With typical D/A converter speeds of something less than a microsecond, you can make sine waves up to a few tens of kilohertz or so, assuming you use angle steps of a degree or so. D/A converters also tend to generate large output spikes ("glitches") while jumping between

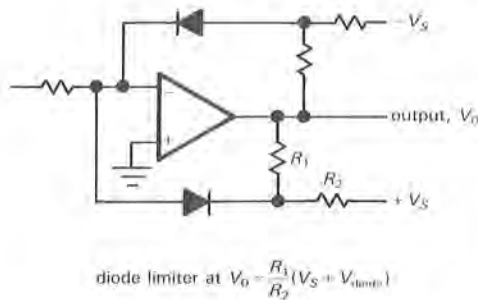


Figure 5.40

output voltages. You can get full-scale glitches even when jumping between adjacent (closest) output voltage levels! In Chapter 9 we'll see deglitching techniques to eliminate this problem. D/A converters are available with resolutions up to 16 bits (1 part in 65,536).

□ State-variable oscillator

The preceding methods all require some hard work. Luckily, the friendly folks at Burr-Brown have done their homework and have come up with the model 4423 "precision quadrature oscillator." It uses the standard 3-op-amp state-variable band-pass filter circuit (Figure 5.18), with the output diode-limited and fed back as input (see Fig. 5.40). It claims to operate from 0.002Hz to 20kHz, with good control of phase shift, amplitude, and frequency

stability (100ppm/°C, max). The 4423 is a module (not a monolithic IC) in a 14-pin molded DIP; it costs \$24 in small quantities.

□ Phase sequence filters

There are tricky *RC* filter circuits that have the property of accepting an input sine wave and producing as output a pair of sine-wave outputs whose phase difference is approximately 90°. The radio hams know this as the "phasing" method of single-sideband generation (due to Weaver), in which the input signal consists of the speech waveform that you want to transmit.

Unfortunately, this method works satisfactorily only over a rather limited range of frequencies and requires precision resistors and capacitors.

A better method for wideband quadrature generation uses "phase sequence networks," consisting of a cyclic repetitive structure of equal resistors and geometrically decreasing capacitors, as in Figure 5.41. You drive the network with a signal and its 180°-shifted cousin (that's easy, since all you need is a unity-gain inverter). The output is a fourfold set of quadrature signals, with a 6-section network giving ±0.5° error over a 100:1 frequency range.

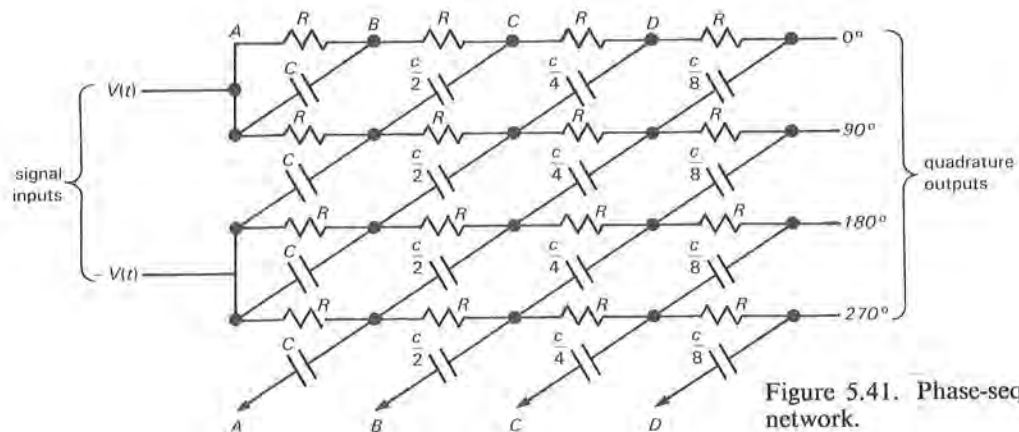


Figure 5.41. Phase-sequence network.

□ Quadrature square waves

For the special case of square waves, generating quadrature signals is a lead-pipe cinch. The basic idea is to generate twice the frequency you need, then divide by 2 with digital *flip-flops* (Chapter 8) and *decode* with *gates* (Chapter 8 again). This technique is essentially perfect from dc to at least 100MHz.

□ Radiofrequency quadrature

At radiofrequencies (upward of a few megahertz) the generation of quadrature sine-wave pairs again becomes easy, using devices known as *quadrature hybrids* (or *quadrature splitter/combiners*). At the low-frequency end of the radio spectrum (from a few megahertz to perhaps 1GHz) these take the form of small core-wound transformers, while at higher frequencies you find incarnations in the form of stripline (strips of foil insulated from an underlying ground plane) or waveguide (hollow rectangular tubing). We'll see these again in Chapter 13. These techniques tend to be fairly narrow-band, with typical operating bandwidths of an octave (i.e., ratio of 2:1).

□ Generating a sine wave of arbitrary phase

Once you have a quadrature pair, it's easy to make a sine wave of *arbitrary* phase. You simply combine the in-phase (*I*) and quadrature (*Q*) signals in a resistive combiner, made most easily with a potentiometer going between the *I* and *Q* signals. As you rotate the pot, you combine the *I* and *Q* in different proportions, taking you smoothly from 0° to 90° phase. If you think in terms of phasors, you'll see that the resulting phase is completely independent of frequency; however, the amplitude varies somewhat as you adjust the phase, dropping 3dB at 45°. You can extend this simple method to the full 360° by simply

generating the inverted (180°-shifted) signals, *I'* and *Q'*, with an inverting amplifier of gain $G_V = -1$.

□ 5.17 Wien bridge and LC oscillators

When a low-distortion sine wave is required, none of the preceding methods is generally adequate. Although wide-range function generators do use the technique of "corrupting" a triangle wave with diode clamps, the resulting distortion can rarely be reduced below 1%. By comparison, most hi-fi audiophiles insist on distortion levels below 0.1% for their amplifiers. To test such low-distortion audio components, pure sine-wave signal sources with residual distortion less than 0.05% or so are required.

At low to moderate frequencies the Wien bridge oscillator (Fig. 5.42) is a good source of low-distortion sinusoidal signals. The idea is to make a feedback amplifier with 180° phase shift at the desired output frequency, then adjust the loop gain so that a self-sustaining oscillation just barely takes place. For equal-value *R*s and *C*s as shown, the voltage gain from the non-inverting input to op-amp output should be exactly +3.00. With less gain the oscillation will cease, and with more gain the output will saturate. The distortion is low if the amplitude of oscillation remains within the linear region of the amplifier, i.e., it must not be allowed to go into a full-swing oscillation. Without some trick to control the gain, that is exactly what will happen, with the amplifier's output increasing until the effective gain is reduced to 3.0 because of saturation. The tricks involve some sort of long-time-constant gain-setting feedback, as you will see.

In the first circuit, an incandescent lamp is used as a variable-resistance feedback element. As the output level rises, the lamp heats slightly, reducing the noninverting gain. The circuit shown has less

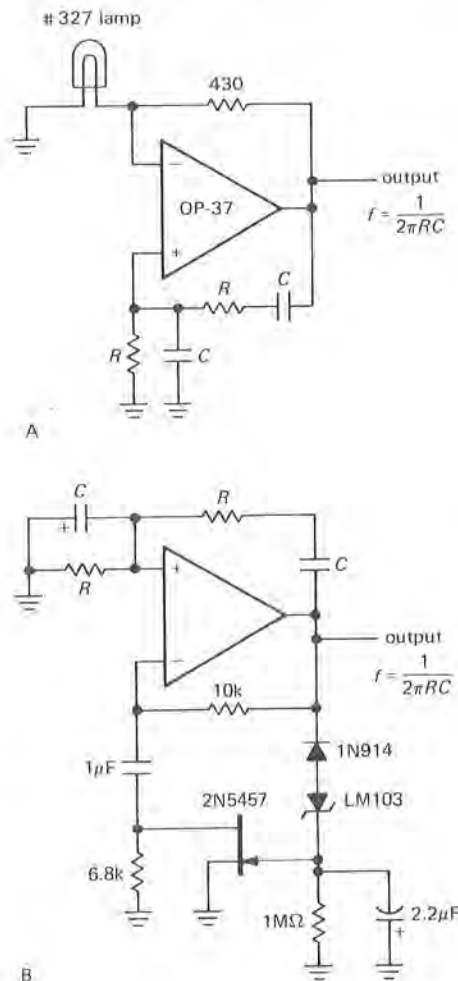


Figure 5.42. Wien-bridge low-distortion oscillators.

than 0.003% harmonic distortion for audiofrequencies above 1kHz; see LTC App. Note 5(12/84) for more details. In the second circuit, an amplitude discriminator consisting of the diodes and RC adjusts the ac gain by varying the resistance of the FET, which behaves like a voltage-variable resistance for small applied voltages (see Section 3.10). Note the long time constant used (2s); this is essential to avoid distortion, since fast feedback will distort the wave by attempting to control the amplitude within the time of one cycle.

□ 5.18 LC oscillators

At high frequencies the favorite method of sine-wave generation is an LC -controlled oscillator, in which a tuned LC is connected in an amplifier-like circuit to provide gain at its resonant frequency. Overall positive feedback is then used to cause a sustained oscillation to build up at the LC 's resonant frequency; such circuits are self-starting.

Figure 5.43 shows two popular configurations. The first circuit is the trusty Colpitts oscillator, a parallel tuned LC at the input, with positive feedback from the output. For this circuit it is claimed that its distortion is less than -60dB . The second circuit is a Hartley oscillator, built with an $n\text{pn}$ transistor. The variable capacitor is for frequency adjustment. Both circuits use *link coupling*, just a few turns of wire acting as a step-down transformer.

LC oscillators can be made *electrically* tunable over a modest range of frequency. The trick is to use a voltage-variable capacitor ("varactor") in the frequency-determining LC circuit. The physics of diode junctions provides the solution, in the form of a simple reverse-biased diode: The capacitance of a pn junction decreases with increasing reverse voltage (see Fig. 13.3). Although any diode acts as a varactor, you can get special varactor diodes designed for the purpose; Figure 5.44 shows some representative types. Figure 5.45 shows a simple JFET Colpitts oscillator (with feedback from the source) with $\pm 1\%$ tunability. In this circuit the tuning range has been made deliberately small, in order to achieve good stability, by using a relatively large fixed capacitor (100pF) shunted by a small tunable capacitor (maximum value of 15pF). Note the large biasing resistor (so the diode bias circuit doesn't load the oscillation) and the dc blocking capacitor. See also Section 13.11.

Varactors typically provide a maximum capacitance of a few picofarads to a few

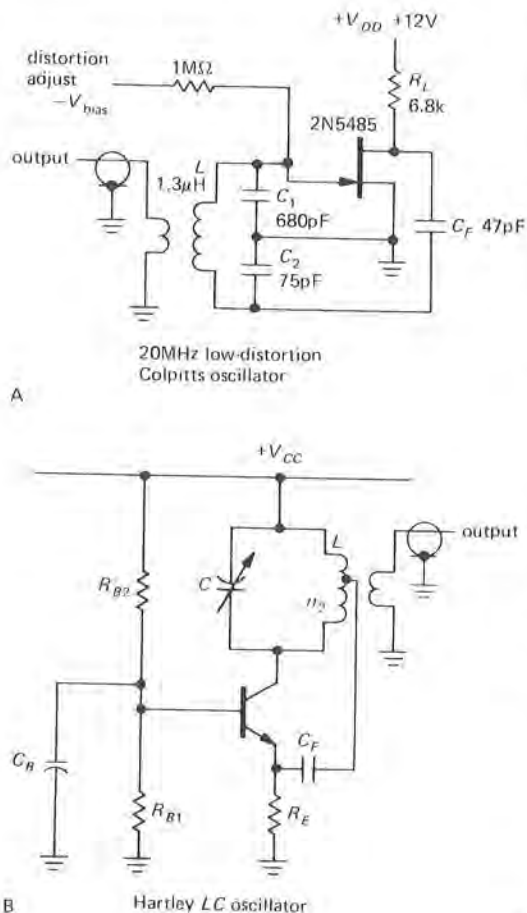


Figure 5.43

hundred picofarads, with a tuning range of about 3:1 (although there are wide range varactors with ratios as high as 15:1). Since the resonant frequency of an LC circuit is inversely proportional to the square root of capacitance, it is possible to achieve tuning ranges of up to 4:1 in frequency, though more typically you're talking about a tuning range of $\pm 25\%$ or so.

In varactor-tuned circuits the oscillation itself (as well as the externally applied dc tuning bias) appears across the varactor, causing its capacitance to vary at the signal frequency. This produces oscillator waveform distortion, and, more important, it

causes the oscillator frequency to depend somewhat on the amplitude of oscillation. In order to minimize these effects, you should limit the amplitude of the oscillation (amplify in following stages, if you need more output); also, it's best to keep the dc varactor bias voltage above a volt or so, in order to make the oscillating voltage small by comparison.

Electrically tunable oscillators are used extensively to generate frequency modulation, as well as in radiofrequency phase-locked loops. We will treat these subjects in Chapters 9 and 13.

For historical reasons we should mention a close cousin of the LC oscillator, namely the tuning-fork oscillator. It used the high- Q oscillations of a tuning fork as the frequency-determining element of an oscillator, and it found use in low-frequency standards (stability of a few parts per million, if run in a constant-temperature oven) as well as wristwatches. These objects have been superseded by quartz oscillators, which are discussed in the next section.

□ Parasitic oscillations

Suppose you have just made a nice amplifier and are testing it out with a sine-wave input. You switch the input function generator to a square wave, but the output remains a sine wave! You don't have an amplifier; you've got trouble.

Parasitic oscillations aren't normally as blatant as this. They are normally observed as fuzziness on part of a waveform, erratic current-source operation, unexplained op-amp offsets, or circuits that behave normally with the oscilloscope probe applied, but go wild when the scope isn't looking. These are bizarre manifestations of untamed high-frequency parasitic oscillations caused by unintended Hartley or Colpitts oscillators employing lead inductance and interelectrode capacitances.

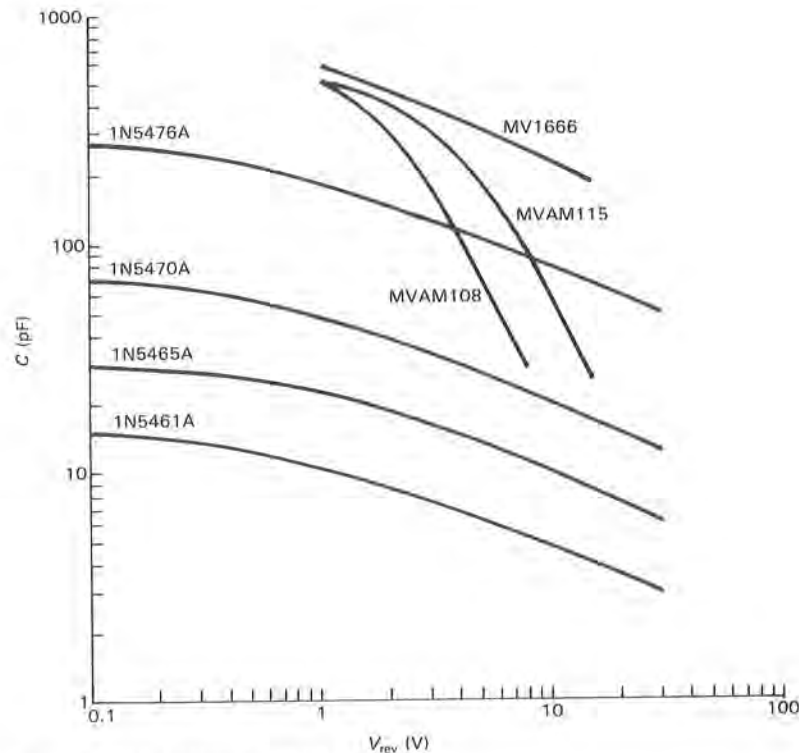


Figure 5.44. Varactor tuning diodes.

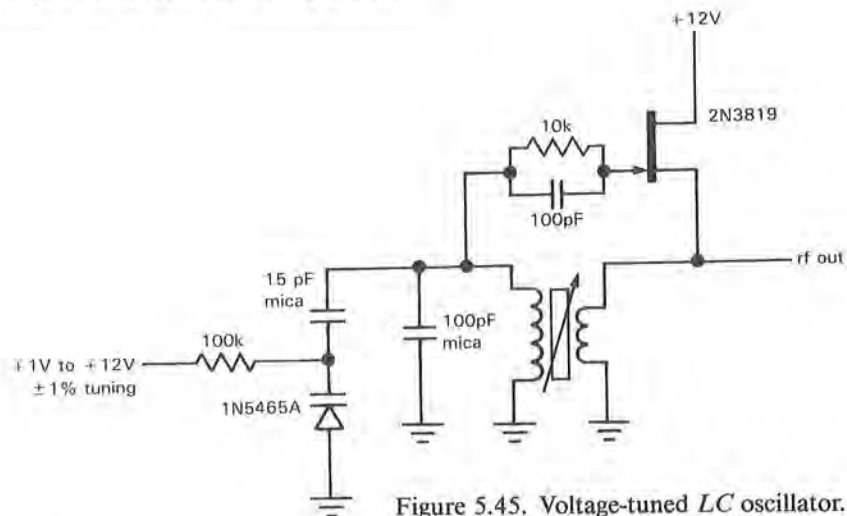


Figure 5.45. Voltage-tuned *LC* oscillator.

The circuit in Figure 5.46 shows an oscillating current source born in an electronics lab course where a VOM was used to measure the output compliance of a standard transistor current source. The current

seemed to vary excessively (5% to 10%) with load voltage variations within its expected compliance range, a symptom that could be “cured” by sticking a finger on the collector lead! The collector-base

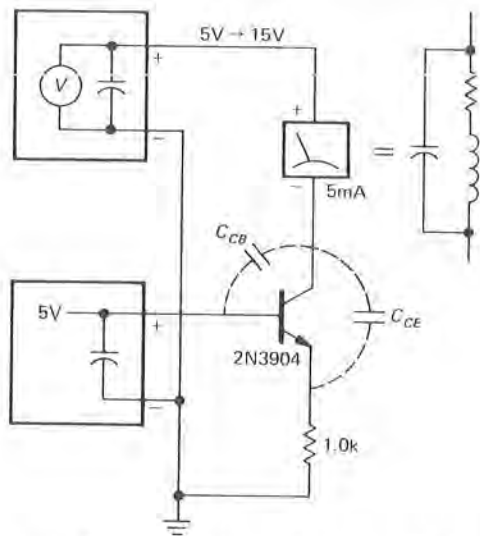


Figure 5.46. Parasitic oscillation example.

capacitance of the transistor and the meter capacitance resonated with the meter inductance in a classic Hartley oscillator circuit, with feedback provided by collector-emitter capacitance. Adding a small base resistor suppressed the oscillation by reducing the high-frequency common-base gain. This is one trick that often helps.

5.19 Quartz-crystal oscillators

RC oscillators can easily attain stabilities approaching 0.1%, with initial predictability of 5% to 10%. That's good enough for many applications, such as the *multiplexed display* in a pocket calculator, in which a multidigit numerical display is driven by lighting one digit after another in rapid succession (a 1kHz rate is typical). Only one digit is lit at any time, but your eye sees the whole display. In such an application the precise rate is quite irrelevant – you just want something in the ballpark. As stable sources of frequency, *LC* oscillators can do a bit better, with stabilities of 0.01% over reasonable periods of time. That's good enough for oscillators in radio-frequency receivers and television sets.

For real stability there's no substitute

for a crystal oscillator. This uses a piece of quartz (same chemical as glass, silicon dioxide) that is cut and polished to vibrate at a certain frequency. Quartz is *piezo-electric* (a strain generates a voltage, and vice versa), so acoustic waves in the crystal can be driven by an applied electric field and in turn can generate a voltage at the surface of the crystal. By plating some contacts on the surface, you wind up with an honest circuit element that can be modeled by an *RLC* circuit, pretuned to some frequency. In fact, its equivalent circuit contains two capacitors, giving a pair of closely spaced (within 1%) series and parallel resonant frequencies (Fig. 5.47). The effect is to produce a rapidly changing reactance with frequency (Fig. 5.48). The quartz crystal's high *Q* (typically around 10,000) and good stability make it a natural for oscillator control, as well as for high-performance filters (see Section 13.12). As with *LC* oscillators, the crystal's equivalent circuit provides positive feedback and gain at the resonant frequency, leading to sustained oscillations.

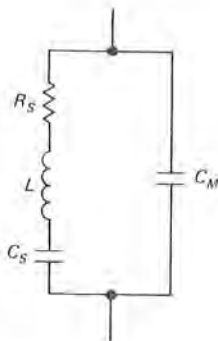


Figure 5.47

Figure 5.49 shows some crystal oscillator circuits. In A the classic Pierce oscillator is shown, using the versatile FET (see Chapter 3). The Colpitts oscillator, with a crystal instead of an *LC*, is shown in B. An *npn* bipolar transistor with the crystal

as feedback element is used in C. The remaining circuits generate logic-level outputs using digital logic functions (D and E).

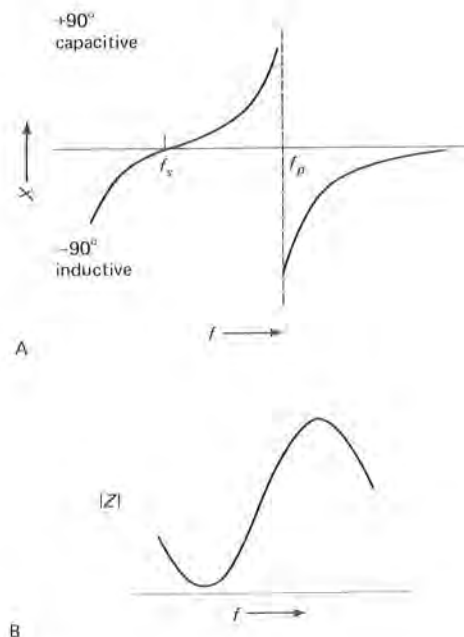


Figure 5.48

The last diagram uses the convenient MC12060/12061 series of crystal oscillator circuits from Motorola. These chips are intended for crystals in the range 100kHz to 20MHz and are designed to give excellent frequency stability by carefully limiting the amplitude of oscillation via internal amplitude discrimination and limiting circuitry. They provide sine-wave and square-wave outputs (both “TTL” and “ECL” logic levels).

An even more convenient alternative, if you’re willing to accept a square wave output only, and if utmost stability isn’t needed, is the use of complete crystal oscillator modules, usually provided as DIP IC-sized metal packages. They come in lots of standard frequencies (e.g., 1, 2, 4, 5, 6, 8, 10, 16, and 20MHz), as well as

weird frequencies commonly used in microprocessor systems (e.g., 14.31818MHz, used for video boards). These “crystal clock modules” typically provide accuracies (over temperature, power supply voltage, and time) of only 0.01% (100ppm), but you get it cheap (\$2 to \$5), and you don’t have to wire up any circuitry. Furthermore, they are guaranteed to oscillate, which isn’t by any means assured when you wire your own oscillator. Crystal oscillator circuits depend on electrical properties of the crystal (such as series versus parallel mode, effective series resistance, and mount capacitance) that aren’t always well specified. All too often you may find that your home-built crystal oscillator oscillates, but at a frequency unrelated to that stamped on the crystal! Our own experience with discrete crystal oscillator circuits has been, well, checkered.

Quartz crystals are available from about 10kHz to about 10MHz, with overtone-mode crystals going to about 250MHz. Although crystals have to be ordered for a given frequency, most of the commonly used frequencies are available off the shelf. Frequencies such as 100kHz, 1.0MHz, 2.0MHz, 4.0MHz, 5.0MHz, and 10.0MHz are always easy to get. A 3.579545MHz crystal (available for less than a dollar) is used in TV color-burst oscillators. Digital wristwatches use 32,768kHz (divide by 2^{15} to get 1Hz), and other powers of 2 are also common. A crystal oscillator can be adjusted slightly by varying a series or parallel capacitor, as shown in Figure 5.49D. Given the low cost of crystals (typically about 2 to 5 dollars), it is worth considering a crystal oscillator in any application where you would have to strain the capabilities of RC relaxation oscillators.

If you need a stable frequency with a very small amount of electrical tunability, you can use a varactor to “pull” the frequency of a quartz-crystal oscillator. The resulting circuit is called a “VCXO”

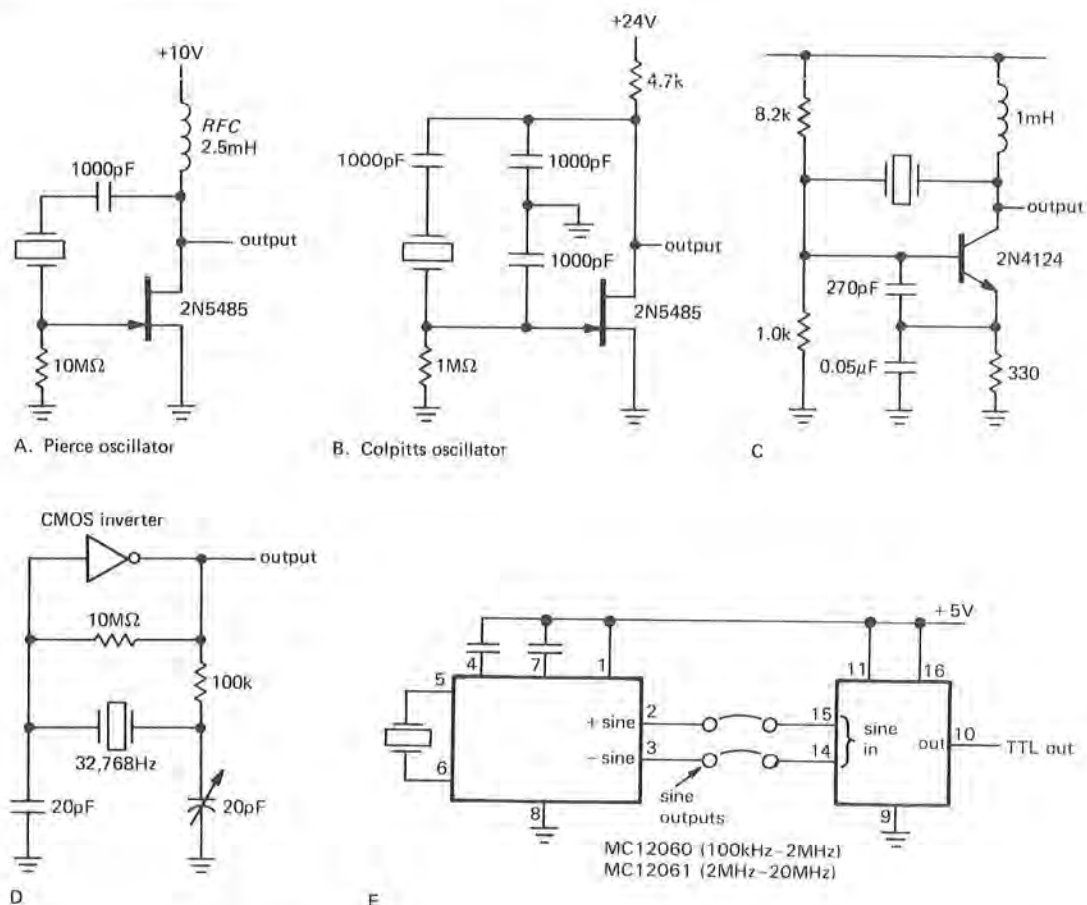


Figure 5.49. Various crystal oscillators.

(voltage-controlled crystal oscillator), and combines the good-to-excellent stability of crystal oscillators with the tunability of *LC* oscillators. The best approach is probably to buy a commercial VCXO, rather than attempt to design your own. Typically they produce maximum deviations of $\pm 10\text{ppm}$ to $\pm 100\text{ppm}$ from center frequency, though wide-deviation units (up to $\pm 1000\text{ppm}$) are also available.

Without great care you can obtain frequency stabilities of a few parts per million over normal temperature ranges with crystal oscillators. By using temperature-compensation schemes you can make a TCXO (temperature-compensated crystal

oscillator) with somewhat better performance. Both TCXOs and uncompensated oscillators are available as complete modules from many manufacturers, e.g., Bliley, CTS Knights, Motorola, Reeves Hoffman, Statek, and Vectron. They come in various sizes, ranging down to DIP packages and TO-5 standard transistor cans. TCXOs deliver stabilities of 1ppm over the range 0°C to 50°C (inexpensive) down to 0.1ppm over the same range (expensive).

Temperature-stabilized oscillators

For the utmost in stability, you may need a crystal oscillator in a constant-temperature

oven. A crystal with a zero temperature coefficient at some elevated temperature (80°C to 90°C) is used, with the thermostat set to maintain that temperature. Such oscillators are available as small modules for inclusion into an instrument or as complete frequency standards ready for rack mounting. The 10811 from Hewlett-Packard is typical of high-performance modular oscillators, delivering 10MHz with stabilities of a few parts in 10^{11} over periods of seconds to hours.

When thermal instabilities have been reduced to this level, the dominant effects become crystal "aging" (the frequency tends to decrease continuously with time), power-supply variations, and environmental influences such as shock and vibration (the latter are the most serious problems in quartz wristwatch design). To give an idea of the aging problem, the oscillator mentioned previously has a specified aging rate at delivery of 5 parts in 10^{10} per day, maximum. Aging effects are due in part to the gradual relief of strains, and they tend to settle down after a few months, particularly in a well-manufactured crystal. Our specimen of the 10811 oscillator ages about 1 part in 10^{11} per day.

Atomic frequency standards are used where the stability of ovenized-crystal standards is insufficient. These use a microwave absorption line in a rubidium gas cell, or atomic transitions in an atomic cesium beam, as the reference to which a quartz crystal is stabilized. Accuracy and stability of a few parts in 10^{12} can be obtained. Cesium-beam standards are the official timekeepers in this country, with timing transmissions from the National Bureau of Standards and the Naval Observatory. Atomic hydrogen masers have been suggested as the ultimate in stable clocks, with claimed stabilities approaching a few parts in 10^{14} . Recent research in stable clocks has centered on techniques using "cooled ions" to achieve

even better stability. Many physicists believe that ultimate stabilities of parts in 10^{18} may be possible.

SELF-EXPLANATORY CIRCUITS

5.20 Circuit ideas

Figure 5.51 presents a variety of circuit ideas, mostly taken from manufacturers' data sheets and applications literature.

ADDITIONAL EXERCISES

1. Design a 6-pole high-pass Bessel filter with cutoff frequency 1kHz.
2. Design a 60Hz twin-T notch filter with op-amp input and output buffers.
3. Design a sawtooth-wave oscillator, to deliver 1kHz, by replacing the charging resistor in the 555 oscillator circuit with a transistor current source. Be sure to provide enough current-source compliance. What value should R_B (Fig. 5.33) have?
4. Make a triangle-wave oscillator with a 555. Use a pair of current sources I_0 (sourcing) and $2I_0$ (sinking). Use the 555's output to switch the $2I_0$ current sink on and off appropriately. The following figure shows one possibility.

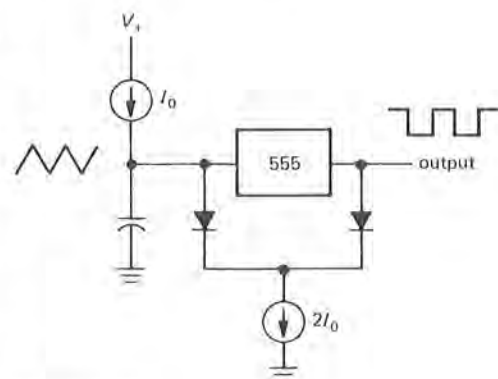
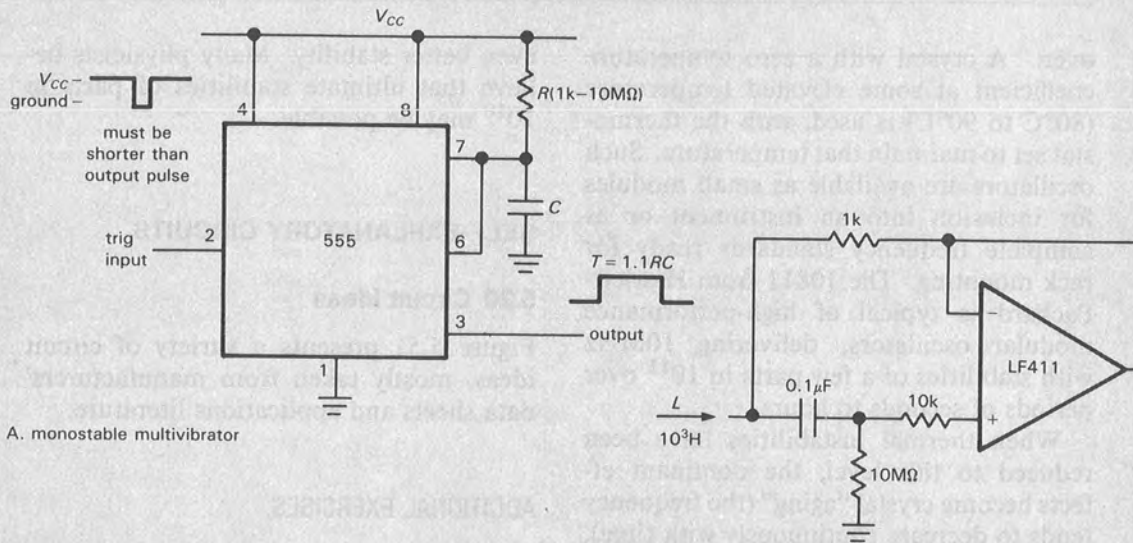


Figure 5.50

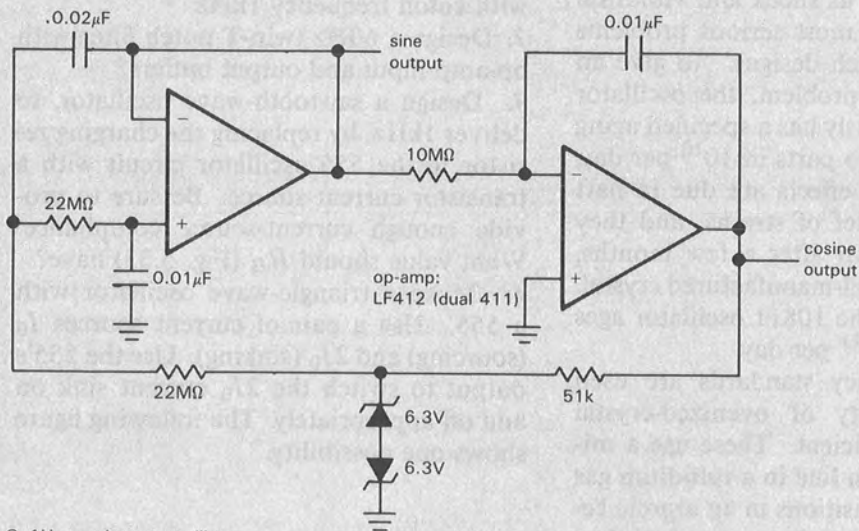


Circuit ideas

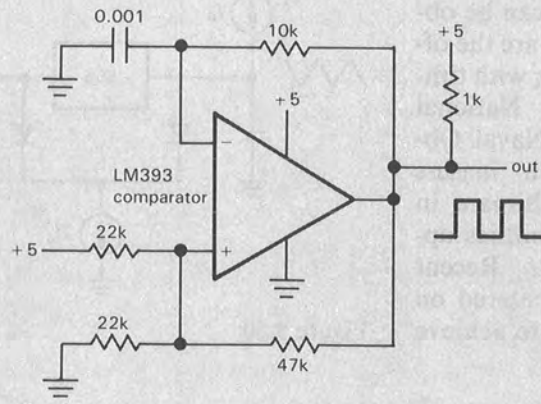


A. monostable multivibrator

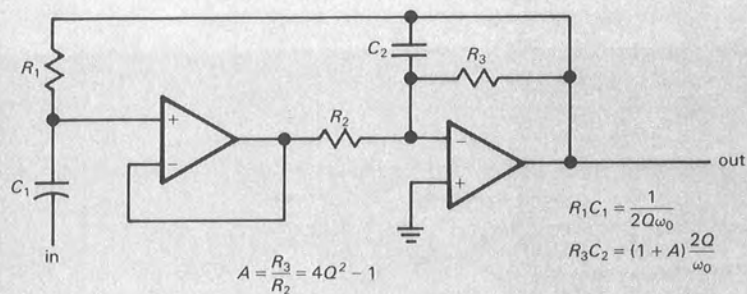
B. active inductor



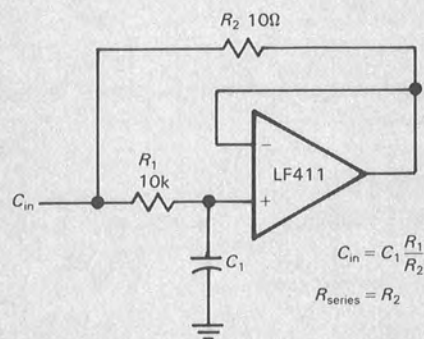
C. 1Hz quadrature oscillator



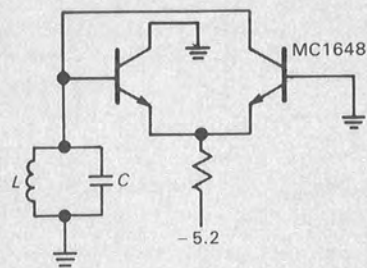
D. relaxation oscillator



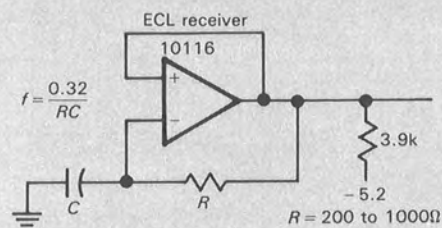
E. tuned amplifier; works for f_0 up to $f_T/2Q$. *Proc. IEEE* 60, 908 (1972)



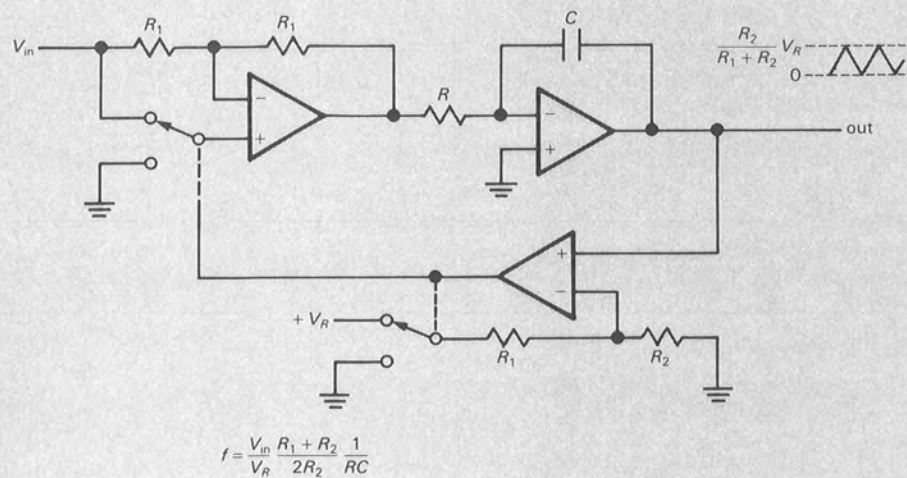
F. capacitance multiplier



G. emitter-coupled LC oscillator



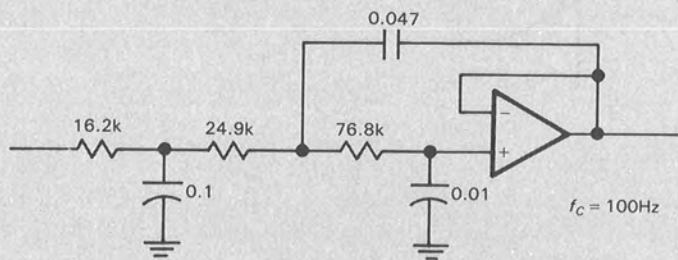
H. high-frequency ECL multivibrator



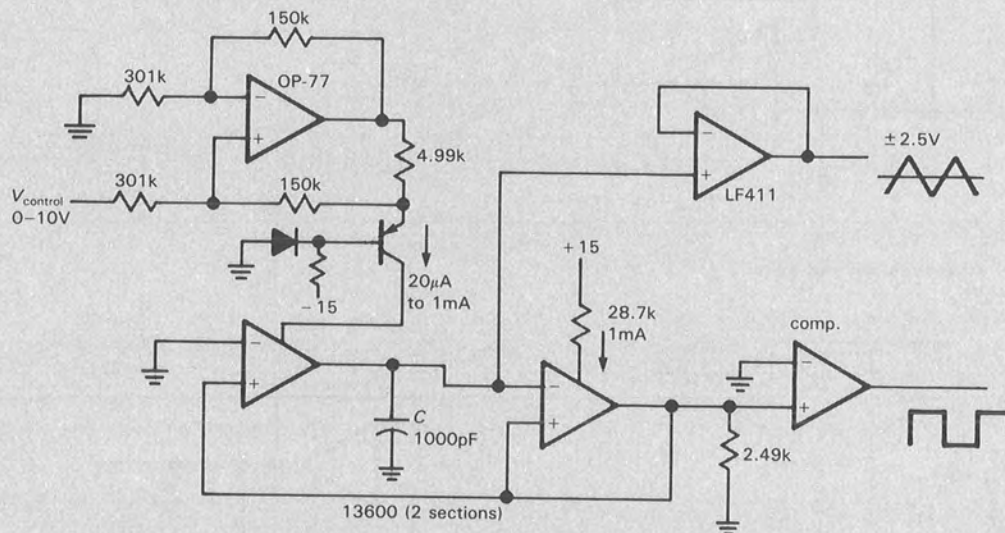
I. voltage-to-frequency converter



Circuit ideas (cont.)



J. 3rd-order Bessel low-pass filter
scale values for other frequencies



K. wide-range VCO (2Hz-100kHz) with
transconductance amplifiers

VOLTAGE REGULATORS AND POWER CIRCUITS

CHAPTER 6

Nearly all electronic circuits, from simple transistor and op-amp circuits up to elaborate digital and microprocessor systems, require one or more sources of stable dc voltage. The simple transformer-bridge-capacitor unregulated power supplies we discussed in Chapter 1 are not generally adequate because their output voltages change with load current and line voltage and because they have significant amounts of 120Hz ripple. Fortunately, it is easy to construct stable power supplies using negative feedback to compare the dc output voltage with a stable voltage reference. Such regulated supplies are in universal use and can be simply constructed with integrated circuit voltage regulator chips, requiring only a source of unregulated dc input (from a transformer-rectifier-capacitor combination, a battery, or some other source of dc input) and a few other components.

In this chapter you will see how to construct voltage regulators using special-purpose integrated circuits. The same circuit techniques can be used to make regulators with discrete components (transistors, resistors, etc.), but because of the availability

of inexpensive high-performance regulator chips, there is no advantage to using discrete components in new designs. Voltage regulators get us into the domain of high power dissipation, so we will be talking about heat sinking and techniques like “foldback limiting” to limit transistor operating temperatures and prevent circuit damage. These techniques can be used for all sorts of power circuits, including power amplifiers. With the knowledge of regulators you will have at that point, we will be able to go back and discuss the design of the unregulated supply in some detail. In this chapter we will also look at voltage references and voltage-reference ICs, devices with uses outside of power-supply design.

BASIC REGULATOR CIRCUITS WITH THE CLASSIC 723

6.01 The 723 regulator

The μ A723 voltage regulator is a classic. Designed by Bob Widlar and first introduced in 1967, it is a flexible, easy-to-use regulator with excellent performance.

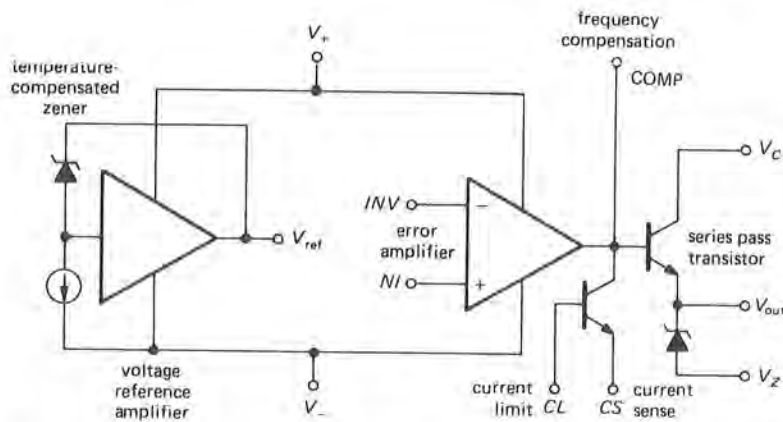


Figure 6.1. Simplified circuit of the 723 regulator. (Courtesy of Fairchild Camera and Instrument Corp.)

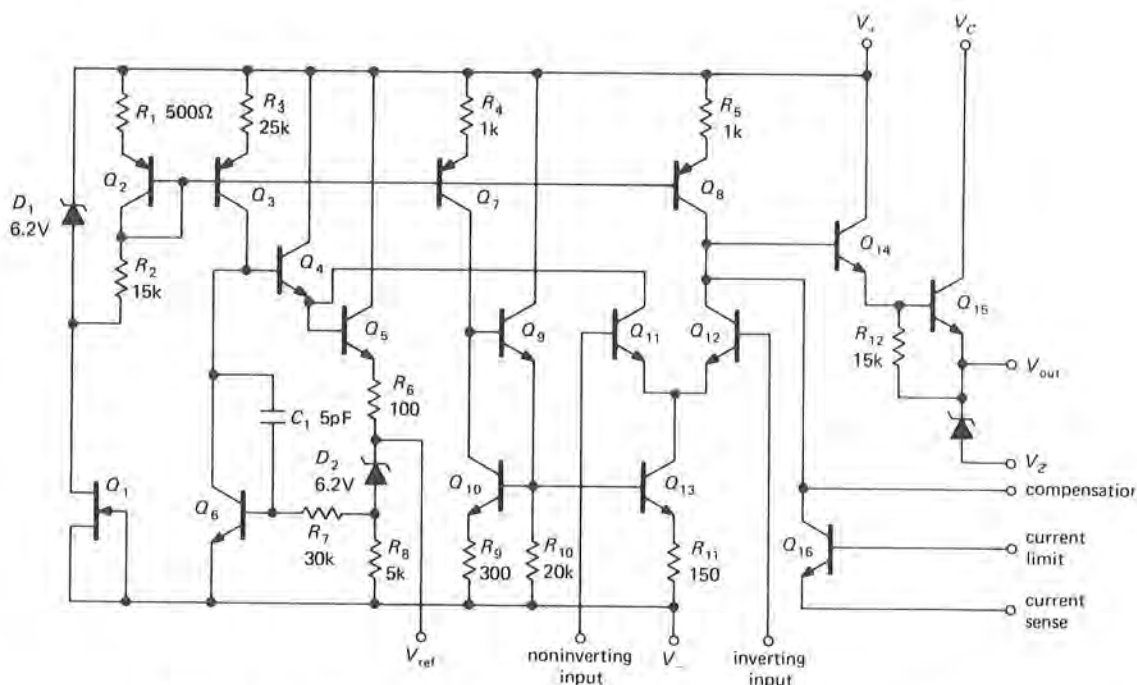


Figure 6.2. Schematic of the 723 regulator. (Courtesy of Fairchild Camera and Instrument Corp.)

Although you would not choose it for a new design nowadays, it is worth looking at in some detail, since more recent regulators work on the same principles. Its circuit is shown in Figures 6.1 and 6.2. As you can see, it is really a power-supply

kit, containing a temperature-compensated voltage reference, differential amplifier, series pass transistor, and current-limiting protective circuit. As it comes, the 723 doesn't regulate anything. You have to hook up an external circuit to make it do

what you want. Before going on to design regulators with it, let's look briefly at its internal circuit. It is straightforward and easy to understand (the innards of many ICs aren't).

The heart of the regulator is the temperature-compensated zener reference. Zener D_2 has a positive temperature coefficient, so its voltage is added to Q_6 's base-emitter drop (remember, V_{BE} has a negative temperature coefficient of roughly $-2\text{mV}/^\circ\text{C}$) to form a voltage reference (nominally 7.15V) of nearly zero temperature coefficient (typically $0.003\%/^\circ\text{C}$). Q_4 through Q_6 are arranged to bias D_2 at $I = V_{BE}/R_8$ via negative feedback at dc, as indicated on the block diagram. Q_2 and Q_3 form an unsymmetrical current mirror to bias the reference; current to the mirror is set by D_1 and R_2 (their junction is fixed at 6.2V below V_+), which in turn is biased by Q_1 (the FET behaves roughly like a current source).

Q_{11} and Q_{12} form the differential amplifier (sometimes called the "error amplifier," thinking of the whole thing as an exercise in negative feedback), a classic long-tailed pair with emitter current source Q_{13} . The latter is half of a current mirror (Q_9 , Q_{10} , and Q_{13}), driven in turn from current mirror Q_7 (Q_3 , Q_7 , and Q_8 all mirror the current generated by the D_1 reference, as we mentioned in Section 2.14). Q_{11} 's collector is tied to the fixed positive voltage at Q_4 's emitter, and the error amplifier's output is taken from Q_{12} 's collector. Current mirror Q_8 supplies the latter's collector load. Q_{14} drives the pass transistor Q_{15} , in a not-quite-Darlington connection. Note that Q_{15} 's collector is brought out separately, to allow for separate positive supplies. By turning on Q_{16} you cut off drive to the pass transistors; this is used to limit output currents to nondestructive levels. Unlike many of the newer regulators, the 723 does not incorporate internal shut-down circuitry to protect against excessive load current or chip

dissipation. The SG3532 and LAS1000 are improved 723-type regulators, with low-voltage bandgap reference (Section 6.15), internal current limiting, and thermal-overload shutdown circuitry.

6.02 Positive regulator

Figure 6.3 shows how to make a positive voltage regulator with the 723. All the components except the four resistors and the two capacitors are contained on the 723. Voltage divider R_1R_2 compares a fraction of the output with the voltage reference, and the 723 components do the rest; this circuit is identical with the op-amp noninverting amplifier with emitter follower, with V_{ref} as the "input." R_4 is chosen for about 0.5 volt drop at maximum desired output current, since a V_{BE} drop applied across the CL-CS inputs will turn on the current-limiting transistor (Q_{16} in Fig. 6.2), shutting off base drive to the output pass transistor. The 100pF capacitor stabilizes the loop. R_3 (sometimes omitted) is chosen so that the differential amplifier sees equal impedances at its inputs. This makes the output insensitive to changes in bias current (with changes in temperature, say), in the same way as we saw with op-amps (Section 4.12).

With this circuit, a regulated supply with output voltage ranging from V_{ref} to the maximum allowable output voltage (37V) can be made. Of course, the input voltage must stay a few volts more positive than the output at all times, including the effects of ripple on the unregulated supply. The "dropout voltage" (the amount by which the input voltage must exceed the regulated output voltage) is specified as 3 volts (minimum) for the 723, a value typical of most regulators. R_1 or R_2 is usually made adjustable, or trimmable, so the output voltage can be set precisely. The production spread in V_{ref} is 6.8 to 7.5 volts.

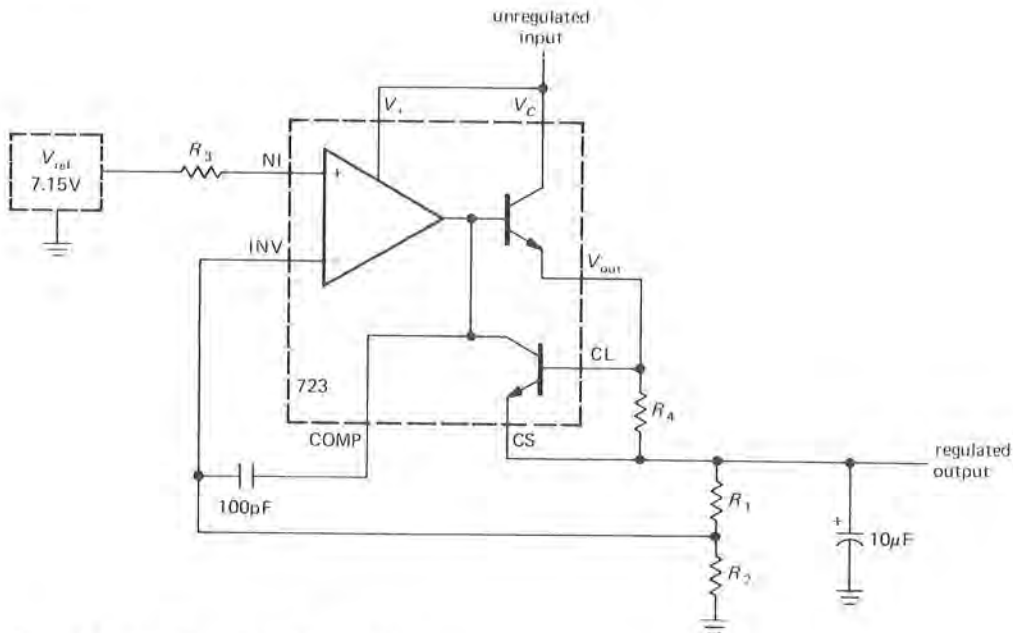


Figure 6.3. 723 regulator: $V_{out} > V_{ref}$.

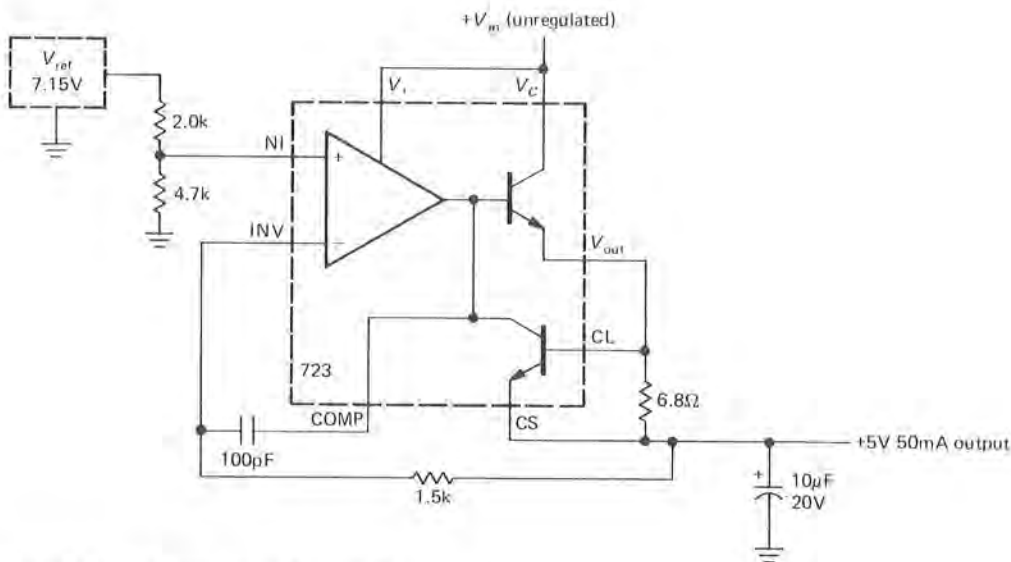


Figure 6.4. 723 regulator: $V_{out} < V_{ref}$.

It is usually a good idea to put a capacitor of a few microfarads across the output, as shown. This keeps the output

impedance low even at high frequencies, where the feedback becomes less effective. It is best to use the output capacitor value

recommended on the specification sheet, since oscillations can occur otherwise. In general, it is a good idea to bypass power-supply leads to ground liberally throughout a circuit, using a combination of ceramic types ($0.01\text{--}0.1\mu\text{F}$) and electrolytic or tantalum types ($1\text{--}10\mu\text{F}$).

For output voltages less than V_{ref} , you just put the voltage divider on the reference (Fig. 6.4). Now the full output voltage is compared with a fraction of the reference. The values shown are for +5 volts 50mA max. With this circuit configuration, output voltages from +2 volts to V_{ref} can be produced. The output cannot be adjusted down to zero volts because the differential amplifier will not operate below 2 volts input. This is given as a manufacturer's specification (see Table 6.9). With this circuit the unregulated input voltage must never drop below +9.5 volts, the voltage necessary to power the reference.

A third variation of this circuit is necessary if you want a regulator that is continuously adjustable through a range of output voltages around V_{ref} . In such cases, just compare a divided fraction of the output with a fraction of V_{ref} chosen to be less than the minimum output voltage desired.

EXERCISE 6.1

Design a regulator to deliver up to 50mA load current over an output voltage range of +5 to +10 volts, using a 723. Hint: Compare a fraction of the output voltage with $0.5V_{\text{ref}}$.

6.03 High-current regulator

The internal pass transistor in the 723 is rated at 150mA maximum; in addition, the power dissipation must not exceed 1 watt at 25°C (less at higher ambient temperatures; the 723 must be "derated" at $8.3\text{mW}/^{\circ}\text{C}$ above 25°C in order to keep the junction temperature within safe limits). Thus, for instance, a 5 volt regulator with +15 volts input cannot

deliver more than about 80mA to the load. To construct a higher-current supply, an external pass transistor must be used. It is easy to add one as a Darlington pair with the internal transistor (Fig. 6.5). Q_1 is the external pass transistor; it must be mounted on a *heat sink*, most often a finned metal plate designed to carry off heat (alternatively, the transistor can be mounted to one wall of the metal chassis housing the power supply). We will deal with thermal problems like these in the next section. A trimmer potentiometer has been used so that the output can be set accurately to +5 volts; its range of adjustment should be sufficient to allow for resistor tolerances as well as the maximum specified spread in V_{ref} (this is an example of worst-case design), and in this case it allows about ± 1 volt adjustment from the nominal output voltage. Note the low-resistance high-power current-limiting resistor necessary for a 2 amp supply.

Pass transistor dropout voltage

One problem with this circuit is the high power dissipation in the pass transistor (at least 10W at full load current). This is unavoidable if the regulator chip is powered by the unregulated input, since it needs a few volts of "headroom" to operate (specified by the dropout voltage). With the use of a separate low-current supply for the 723 (e.g., +12V), the minimum unregulated input to the external pass transistor can be only a volt or so above the regulated output voltage (although you will always have to allow at least a few volts, since worst-case design dictates proper operation even at 105V ac line input).

Overvoltage protection

Also shown in this circuit is an *overvoltage crowbar* protection circuit consisting of

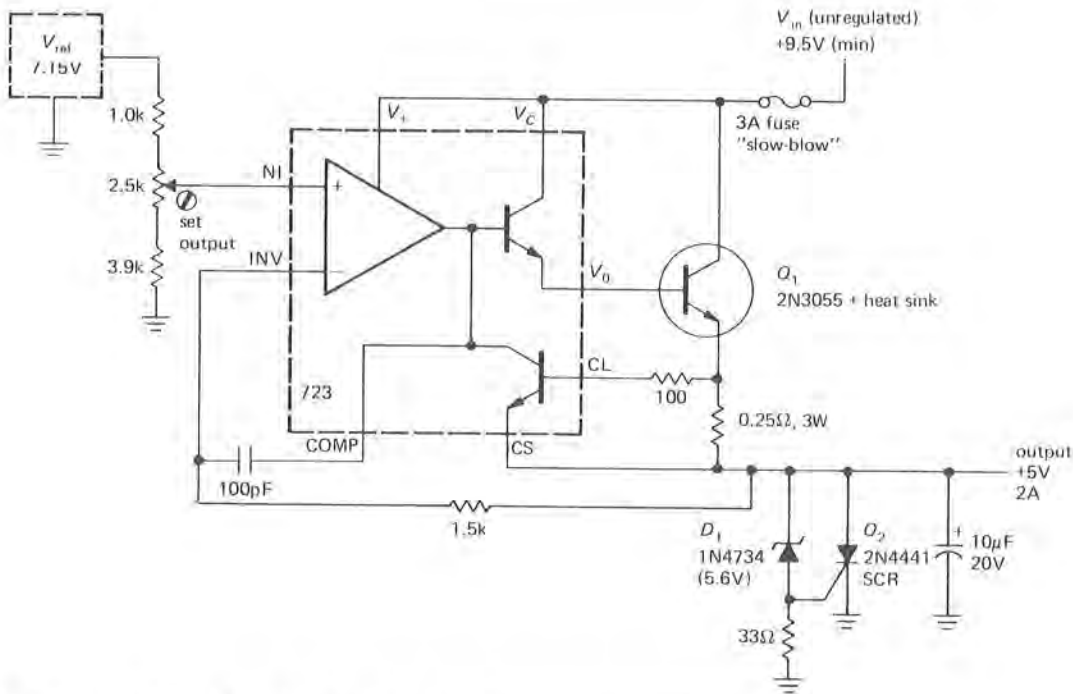


Figure 6.5. Five volt regulator with outboard pass transistor and crowbar.

D_1 , Q_2 , and the 33 ohm resistor. Its function is to short the output if some circuit fault causes the output voltage to exceed about 6.2 volts (this could happen if one of the resistors in the divider were to open up, for instance, or if some component in the 723 were to fail). Q_2 is an SCR (silicon-controlled rectifier), a device that is normally nonconducting but that goes into saturation when the gate-cathode junction is forward-biased. Once turned on, it will not turn off again until anode current is removed externally. In this case, gate current flows when the output exceeds D_1 's zener voltage plus a diode drop. When that happens, the regulator will go into a current-limiting condition, with the output held near ground by the SCR. If the failure that produces the abnormally high output also disables the current-limiting circuit (e.g., a collector-to-emitter short in Q_1), then the crowbar will sink a very large current. For this reason it is a good idea to include a

fuse somewhere in the power supply, as shown. We will treat overvoltage crowbar circuits in more detail in Section 6.06.

HEAT AND POWER DESIGN

6.04 Power transistors and heat sinking

As in the preceding circuit, it is often necessary to use power transistors or other high-current devices like SCRs or power rectifiers that can dissipate many watts. The 2N3055, an inexpensive power transistor of great popularity, can dissipate as much as 115 watts if properly mounted. All power devices are packaged in cases that permit contact between a metal surface and an external heat sink. In most cases the metal surface of the device is electrically connected to one terminal (e.g., for power transistors the case is always connected to the collector).

The whole point of heat sinking is to keep the transistor junction (or the junction of some other device) below some maximum specified operating temperature. For silicon transistors in metal packages the maximum junction temperature is usually 200°C, whereas for transistors in plastic packages it is usually 150°C. Table 6.1 lists some useful power transistors, along with their thermal properties. Heat sink design is then simple: Knowing the maximum power the device will dissipate in a given circuit, you calculate the junction temperature, allowing for the effects of heat conductivity in the transistor, heat sink, etc., and the maximum ambient temperature in which the circuit is expected to operate. You then choose a heat sink large enough to keep the junction temperature well below the maximum specified by the manufacturer. It is wise to be conservative in heat sink design, since transistor life drops rapidly at operating temperatures near or above maximum.

Thermal resistance

To carry out heat sink calculations, you use *thermal resistance*, θ , defined as heat rise (in degrees) divided by power transferred. For heat transferred entirely by conduction, the thermal resistance is a constant, independent of temperature, that depends only on the mechanical properties of the joint. For a succession of thermal joints in "series," the total thermal resistance is the sum of the thermal resistances of the individual joints. Thus, for a transistor mounted on a heat sink, the total thermal resistance from transistor junction to the outside (ambient) world is the sum of the thermal resistance from junction to case θ_{JC} , the thermal resistance from case to heat sink, θ_{CS} , and the thermal resistance from heat sink to ambient θ_{SA} . The temperature of the junction is therefore

$$T_J = T_A + (\theta_{JC} + \theta_{CS} + \theta_{SA})P$$

where P is the power being dissipated.

Let's take an example. The preceding power-supply circuit, with external pass transistor, has a maximum transistor dissipation of 20 watts for an unregulated input of +15 volts (10V drop, 2A). Let's assume that the power supply is to operate at ambient temperatures up to 50°C, not unreasonable for electronic equipment packaged together in close quarters. And let's try to keep the junction temperature below 150°C, well below its specified maximum of 200°C. The thermal resistance from junction to case is 1.5°C per watt. A TO-3 power transistor package mounted with an insulating washer and heat-conducting compound has a thermal resistance from case to heat sink of about 0.3°C per watt. Finally, a Wakefield model 641 heat sink (Fig. 6.6) has a thermal resistance from sink to ambient of about 2.3°C per watt. So the total thermal resistance from junction to ambient is about 4.1°C per watt. At 20 watts dissipation the junction will be 84°C above ambient, or 134°C (at maximum ambient temperature) in this example. The chosen heat sink will be adequate; in fact, a smaller one could be used if necessary to save space.

Comments on heat sinks

1. Where very high power dissipation (several hundred watts, say) is involved, forced air cooling may be necessary. Large heat sinks designed to be used with a blower are available with thermal resistances (sink to ambient) as small as 0.05°C to 0.2°C per watt.
2. When the transistor must be insulated from the heat sink, as is usually necessary (especially if several transistors are mounted on the same sink), a thin insulating washer is used between the transistor and sink, and insulating bushings are used around the mounting screws. Washers are available in standard

TABLE 6.1. SELECTED BIPOLAR POWER TRANSISTORS

npn	pnP	Pkg ^a	V _{CE0} max (V)	I _C max (A)	h _{FE} typ	@ I _C (A)	f _T min (MHz)	C _{cb} ^b typ (pF)	P _{diss} (T _C =25°C) (W)	Θ _{JC} (°C/W)	T _J max (°C)	Comments
Regular power: V _{CE(sat)} = 0.4V (typ); V _{BE(on)} = 0.8V (typ)												
2N5191	2N5194	A	60	4	100	0.2	2	80	40	3.1	150	low cost, gen purp
2N5979	2N5976	B	80	5	50	0.5	2	60	70	1.8	150	
2N3055	MJ2955	TO-3	60	15	50	2	2.5	125	115	1.5	200	metal, indus std
MJE3055	MJE2955	B	60	10	50	2	2.5	125	90	1.4	150	plastic, indus std
2N5886	2N5884	TO-3	80	25	50	10	4	400	200	0.9	200	
2N5686	2N5684	TO-3	80	50	30	25	2	700	300	0.6	200	for real power job
2N6338	2N6437	TO-3	100	25	50	8	40	200	200	0.9	200	premium audio
2N6275	2N6379	TO-3	120	50	50	20	30	400	250	0.7	200	premium audio
Darlington power: V _{CE(sat)} = 0.8V (typ); V _{BE(on)} = 1.4V (typ)												
2N6038	2N6035	A	60	4	2000	2	—	30	40	3.1	150	low cost
2N6044	2N6041	B	80	8	2500	4	4	80	75	1.7	150	
2N6059	2N6052	TO-3	100	12	3500	5	4	100	150	1.2	200	
2N6284	2N6287	TO-3	100	20	3000	10	4	150	160	1.1	200	high current

(^a) A: small plastic pwr pkg (TO-126). B: large plastic pwr pkg (TO-127). (^b) C_{cb} (npn) at V_{CB}=10V; C_{cb} (pnP) = 2C_{cb} (npn).

transistor-shape cutouts made from mica, insulated aluminum, or beryllia (BeO). Used with heat-conducting grease, these add from 0.14°C per watt (beryllia) to about 0.5°C per watt.

An attractive alternative to the classic mica-washer-plus-grease is provided by greaseless silicone-based insulators that are loaded with a dispersion of thermally conductive compound, usually boron nitride or aluminum oxide. They're clean and dry, and easy to use; you don't get white slimy stuff all over your hands, your electronic device, and your clothes. You save lots of time. They have thermal resistances of about 0.2–0.4°C per watt, comparable to values with the messy method. Bergquist calls its product "Sil-Pad," Chomerics calls its "Cho-Therm," SPC calls it "Koolex," and Thermalloy calls its "Thermasil." We've been using these insulators, and we like them.

3. Small heat sinks are available that simply clip over the small transistor packages (like the standard TO-5). In situations of

relatively low power dissipation (a watt or two) this often suffices, avoiding the nuisance of mounting the transistor remotely on a heat sink with its leads brought back to the circuit. An example is shown in Figure 6.6. In addition, there are various small heat sinks intended for use with the plastic power packages (many regulators, as well as power transistors, come in this package) that mount right on a printed-circuit board underneath the package. These are very handy in situations of a few watts dissipation; a typical unit is illustrated in Figure 6.6.

4. Sometimes it may be convenient to mount power transistors directly to the chassis or case of the instrument. In such cases it is wise to use conservative design (keep it cool), especially since a hot case will subject the other circuit components to high temperatures and shorten component life.

5. If a transistor is mounted to a heat sink without insulating hardware, the heat sink must be insulated from the chassis.

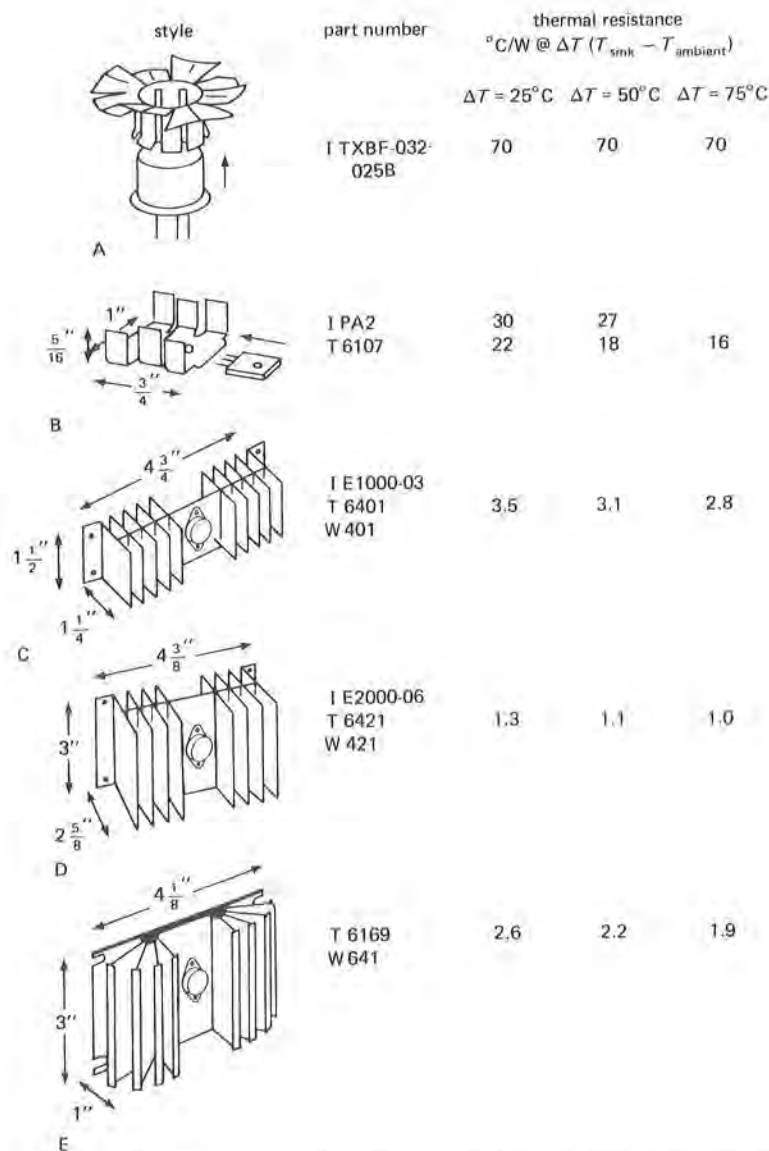


Figure 6.6. Power transistor heat sinks. I, IERC; T, Thermalloy; W, Wakefield.

The use of insulating washers (e.g., Wakefield model 103) is recommended (unless, of course, the transistor case happens to be at ground). When the transistor is insulated from the sink, the heat sink may be attached directly to the chassis. But if the transistor is accessible from outside the instrument (e.g., if the heat sink is mounted

externally on the rear wall of the box), it is a good idea to use an insulating cover over the transistor (e.g., Thermalloy 8903N) to prevent someone from accidentally coming in contact with it, or shorting it to ground.

6. The thermal resistance from heat sink to ambient is usually specified for the sink

mounted with the fins vertical and with unobstructed flow of air. If the sink is mounted differently, or if the air flow is obstructed, the efficiency will be reduced (higher thermal resistance); usually it is best to mount it on the rear of the instrument with fins vertical.

EXERCISE 6.2

A 2N5320, with a thermal resistance from junction to case of 17.5°C per watt, is fitted with an IERC TXBF slip-on heat sink of the type shown in Figure 6.6. The maximum permissible junction temperature is 200°C . How much power can you dissipate with this combination at 25°C ambient temperature? How much must the dissipation be decreased per degree rise in ambient temperature?

□ 6.05 Foldback current limiting

For a regulator with simple current limiting, transistor dissipation is maximum when the output is shorted to ground (either accidentally or through some circuit malfunction), and it usually exceeds the maximum value of dissipation that would otherwise occur under normal load conditions. For instance, the pass transistor in the preceding +5 volt 2 amp regulator circuit will dissipate 30 watts with the output shorted (+15V input, current limit at 2A), whereas the worst-case dissipation under normal load conditions is 20 watts (10V drop at 2A). The situation is even worse in circuits in which the voltage normally dropped by the pass transistor is a smaller fraction of the output voltage. For instance, in a +15 volt 2 amp regulated supply with +25 volt unregulated input, the transistor dissipation rises from 20 watts (full load) to 50 watts (short circuit).

You get into a similar problem with push-pull power amplifiers. Under normal conditions you have maximum load current when the voltage across the transistors is minimum (near the extremes of output swing), and you have maximum voltage

across the transistors when the current is nearly zero (zero output voltage). With a short-circuit load, on the other hand, you have maximum load current at the worst possible time, namely, with full supply voltage across the transistor. This results in much higher transistor dissipation than normal.

The brute-force solution to this problem is to use massive heat sinks and transistors of higher power rating (and safe operating area, see Section 6.07) than necessary. Even so, it isn't a good idea to have large currents flowing into the powered circuit under fault conditions, since other components in the circuit may then be damaged. The best solution is to use *foldback* current limiting, a circuit technique that reduces the output current under short-circuit or overload conditions. Figure 6.7 shows the basic configuration, again illustrated with a 723 with external pass transistor.

The divider at the base of the current-limiting transistor Q_L provides the foldback. At +15 volts output (the normal value) the circuit will limit at about 2 amps, since Q_L 's base is then at +15.5 volts while its emitter is at +15 (V_{BE} is about 0.5V at the elevated temperatures at which regulator chips are normally run). But the short-circuit current is less; with the output shorted to ground, the output current is about 0.5 amp, holding Q_1 's dissipation down to less than in the full-load case. This is highly desirable, since excessive heat sinking is not now required, and the thermal design need only satisfy the full-load requirements. The choice of the three resistors in the current-limiting circuit sets the short-circuit current, for a given full-load current limit. Warning: Use care in choosing the short-circuit current, since it is possible to be overzealous and design a supply that will not "start up" into a normal load. The short-circuit current should not be too small; as a

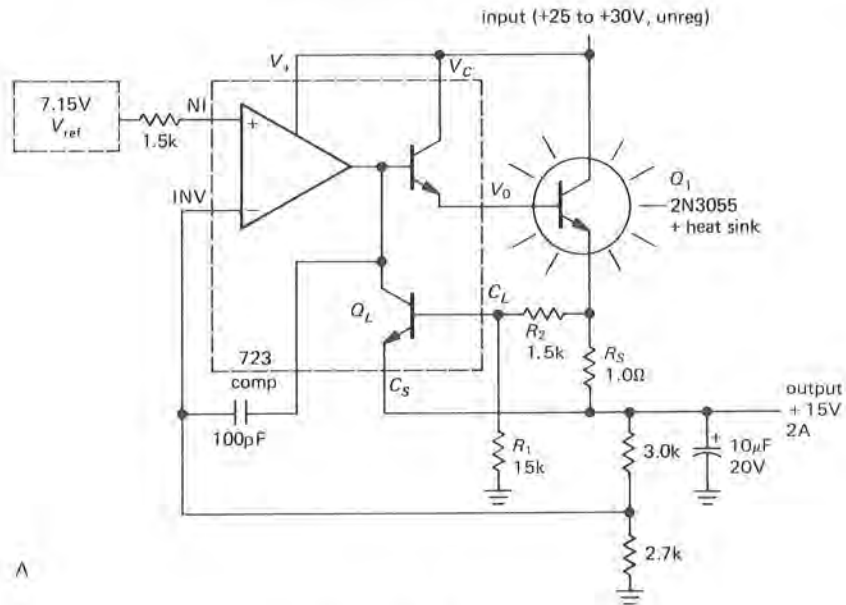
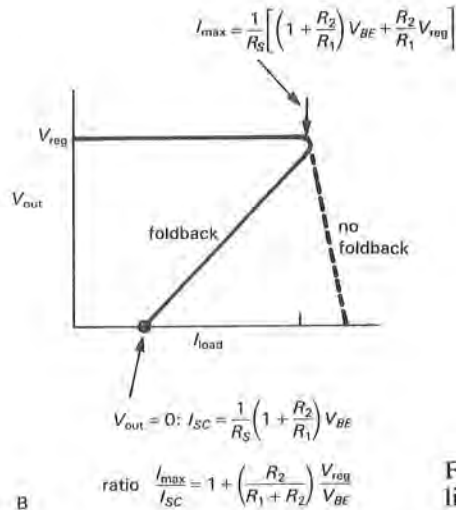


Figure 6.7. A. Power regulator with foldback current limiting. B. Output voltage versus load current.

rough guide, the short-circuit current limit should be set at about one-third the maximum load current at full output voltage.

EXERCISE 6.3

Design a 723 regulator with outboard pass transistor and foldback current limiting to provide up to 1.0 amp when the output is at its regulated value of +5.0 volts, but only 0.4 amp into a short-circuit load.

6.06 Overvoltage crowbars

As we remarked in Section 6.03, it is often a good idea to include some sort of overvoltage protection at the output of a regulated supply. Take, for instance, a +5 volt supply used to power a large digital system (you'll see lots of examples beginning in Chapter 8). The input to the regulator is probably in the range of +10 to +15 volts. If the series pass transistor fails by shorting its collector

to emitter (a common failure mode), the full unregulated voltage will be applied to the circuit, with devastating results. Although a fuse probably will blow, what's involved is a race between the fuse and the "silicon fuse" that is constituted by the rest of the circuit; the rest of the circuit will probably respond first! This problem is most serious with TTL logic, which operates from a +5 volt supply, but cannot tolerate more than +7 volts without damage. Another situation with considerable disaster potential arises when you operate something from a wide-range "bench" supply, where the unregulated input may be 40 volts or more, regardless of the output voltage.

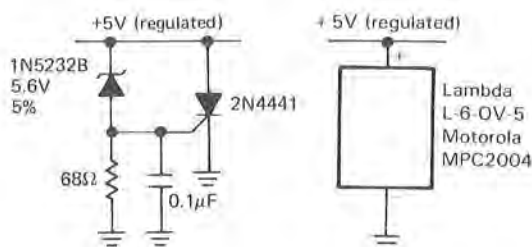


Figure 6.8. Overvoltage crowbars.

□ Zener sensing

Figure 6.8 shows a popular crowbar circuit and a crowbar module. You hook the circuit between the regulated output terminal and ground. If the voltage exceeds the zener voltage plus a diode drop (about 6.2V for the zener shown), the SCR is turned on, and it remains in a conducting state until its anode current drops below a few milliamps. An inexpensive SCR like the 2N4441 can sink 5 amps continuously and withstand 80 amp surge currents; its voltage drop in the conducting state is typically 1.0 volt at 5 amps. The 68 ohm resistor is provided to generate a reasonable zener current (10mA) at SCR turn-on, and the capacitor is added to prevent crowbar triggering on harmless short spikes.

The preceding circuit, like all crowbars, puts an unrelenting 1 volt "short circuit" across the supply when triggered by an overvoltage condition, and it can be reset only by turning off the supply. Since the SCR maintains a low voltage while conducting, there isn't much problem with the crowbar itself failing from overheating. As a result, it is a reliable crowbar circuit. It is essential that the regulated supply have some sort of current limiting, or at least fusing, to handle the short. There may be overheating problems with the supply after the crowbar fires. In particular, if the supply includes internal current limiting, the fuse won't blow, and the supply will sit in the "crowbarred" state, with the output at low voltage, until someone notices. Foldback current limiting of the regulated supply would be a good solution here.

There are several problems with this simple crowbar circuit, mostly involving the choice of zener voltage. Zeners are available in discrete values only, with generally poor tolerances and (often) soft knees in the $V-I$ characteristic. The desired crowbar trigger voltage may involve rather tight tolerances. Consider a 5 volt supply used to power digital logic. There is typically a 5% or 10% tolerance on the supply voltage, meaning that the crowbar cannot be set less than 5.5 volts. The minimum permissible crowbar voltage is raised by the problem of transient response of a regulated supply: When the load current is changed quickly, the voltage can jump, creating a spike followed by some "ringing." This problem is exacerbated by remote sensing via long (inductive) sense leads. The resultant ringing puts glitches on the supply that we don't want to trigger the crowbar. The result is that the crowbar voltage should not be set less than about 6.0 volts, but it cannot exceed 7.0 volts without risk of damage to the logic circuits. When you fold in zener tolerance, the discrete voltages actually available, and SCR trigger voltage tolerances, you've got

a tricky problem. In the example shown earlier, the crowbar threshold could lie between 5.9 volts and 6.6 volts, even using the relatively precise 5% zener indicated.

□ IC sensing

A nice solution to the problems of predictability and lack of adjustability in the simple zener/SCR crowbar circuit is to use a special crowbar trigger IC such as the MC3423-5, the TL431, or the MC34061-2. These inexpensive chips come in convenient packages (8-pin mini-DIP or 3-pin TO-92), they drive the SCR directly, and they're very easy to use. For example, the MC3425 has adjustable threshold and response time for its crowbar output, and in addition an *undervoltage* sensor to signal your circuit that the supply voltage is low (very handy for circuits with microprocessors). It includes an internal reference and several comparators and drivers, and it requires only two external resistors, an optional capacitor, and an SCR to form a complete crowbar. These crowbar chips belong to a class of "power-supply supervisory circuits," which includes complex chips like the MAX691 that not only sense undervoltage but even switch over to battery backup when ac power fails, generate a power-on reset signal on return of normal power, and continually check for lockup conditions in microprocessor circuitry.

Modular crowbars

Why build it when you can buy it! From the designer's point of view the simplest crowbar of all is a 2-terminal gadget that says "crowbar" on top. You can buy just such a device from Lambda or Motorola, who offer a series of overvoltage protection modules in several current ranges. You just pick the voltage and current rating you need, and connect the crowbar across the regulated dc output. For example, the smallest units from Lambda are rated at 2 amps maximum, with the following

set of fixed voltages (5V, 6V, 12V, 15V, 18V, 20V, and 24V). They're monolithic, come in a TO-66 package (small metal power transistor case), and cost \$2.50 in small quantities. The Lambda monolithic 6 amp series comes in TO-3 packages (large metal power transistor case) and costs \$5. They also make hybrid 12, 20, and 35 amp crowbars. Motorola's MPC2000 series are all monolithic (5V, 12V, and 15V only, rated at 7.5A, 15A, or 35A). The first two come in TO-220 (plastic power) packages, the last (available in 5V only) in TO-3 (metal power). The good news from Motorola is the incredibly low price: \$1.96, \$2.36, and \$6.08 in small quantities for the three current ratings. One nice feature of these crowbars is the good accuracy; for example, the 5 volt units from Lambda have a specified trip point of 6.6 ± 0.2 volts.

□ Clamps

Another possible solution to overvoltage protection is to put a power zener, or its equivalent, across the supply terminals. This avoids the problems of false triggering on spikes, since the zener will stop drawing current when the overvoltage condition disappears (unlike an SCR, which has the memory of an elephant). Figure 6.9 shows

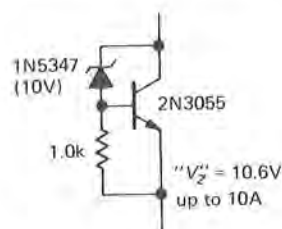


Figure 6.9. Active power zener.

the circuit of an "active zener." Unfortunately, a crowbar constructed from a power zener clamp has its own problems. If the regulator fails, the crowbar has to contend with high power dissipation ($V_{\text{zener}} I_{\text{limit}}$)

and may itself fail. We witnessed just such a failure in a commercial 15 volt 4 amp magnetic disc supply. When the pass transistor failed, the 16 volt 50 watt zener found itself dissipating more than rated power, and it proceeded to fail too.

□ 6.07 Further considerations in high-current power-supply design

□ *Separate high-current unregulated supply*

As we mentioned in Section 6.03, it is usually a good idea to use a separate supply to power the regulator in very high current supplies. In that way the dissipation in the pass transistors can be minimized, since the unregulated input to the pass transistor can then be chosen just high enough to allow sufficient “headroom” (regulators like the 723 have separate V_+ terminals for this purpose). For instance, a +5 volt 10 amp regulator might use a 10 volt unregulated input with a volt or two of ripple, with a separate low-current +15 volt supply for the regulator components (reference, error amplifier, etc.). As mentioned earlier, the unregulated input voltages must be chosen large enough to allow for worst-case ac power-line voltage (105V) as well as transformer and capacitor tolerances.

□ *Connection paths*

With high-current supplies, or supplies of highly precise output voltage, careful thought must be given to the connection paths, both within the regulator and between the regulator and its load. If several loads are run from the same supply, they should connect to the supply at the place where the output voltage is sensed; otherwise, fluctuations in the current of one load will affect the voltage seen by the other loads (Fig. 6.10).

In fact, it is a good idea to have one common ground point (a “mecca”), as

shown, to which the unregulated supply, reference, etc., are all returned. The problem of unregulated voltage drops in the connecting leads from power supply to high-current load is sometimes solved by remote sensing: The connections back to the error amplifier and reference are brought out to the rear of the supply separately and may either be connected to the output terminals right there (the normal method) or brought out and connected to the load at a remote location along with the output voltage leads (this requires four wires, two of which must be able to handle

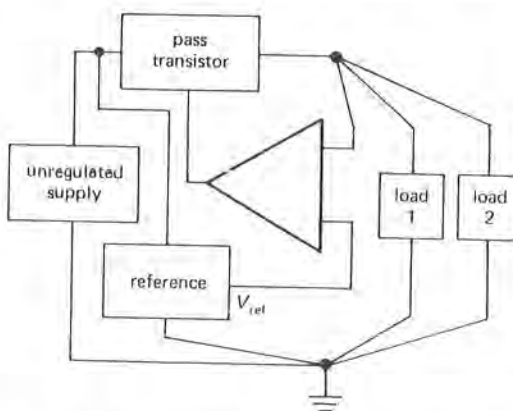


Figure 6.10. A power-supply ground “mecca.”

the high load currents). Most commercially available power supplies come with jumpers at the rear that connect the sensing circuitry to the output and that may be removed for remote sensing. Four-wire resistors are used in an analogous manner to sense load currents accurately when constructing precision constant-current supplies. This will be discussed in greater detail in Section 6.24.

□ *Parallel pass transistors*

When very high output currents are needed, it may be necessary to use several pass transistors in parallel. Since there will be a spread of V_{BE} s, it is necessary to

add a small resistance in series with each emitter, as in Figure 6.11. The R s ensure that the current is shared approximately equally among the pass transistors. R should be chosen for about 0.2 volt drop at maximum output current. Power FETs can be connected in parallel without any external components, owing to their negative temperature coefficient of drain current (Fig. 3.13).

Safe operating area

One last point about bipolar power transistors: A phenomenon known as “second breakdown” restricts the simultaneous

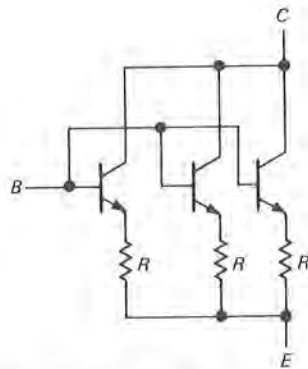


Figure 6.11. Use emitter ballasting resistors when paralleling bipolar power transistors.

voltage and current that may be applied for any given transistor, and it is specified on the data sheet as the safe operating area (SOA) (it's a family of safe voltage-versus-current regions, as a function of time duration). Second breakdown involves the formation of “hot spots” in the transistor junctions, with consequent uneven sharing of the total load. Except at low collector-to-emitter voltages, it sets a limit that is more restrictive than the maximum power dissipation specification. As an example, Figure 6.12 shows the SOA for the ever-popular 2N3055. For $V_{CE} > 40$ volts, second breakdown limits the dc collector current to values corresponding to less

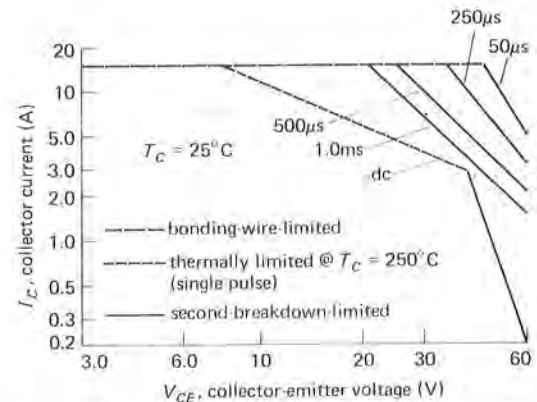


Figure 6.12. Safe operating area for 2N3055 bipolar power transistor. (Courtesy of Motorola, Inc.)

than the maximum allowable dissipation of 115 watts. Figure 6.13 shows the SOA for two similar high-performance power transistors: the 2N6274 *npn* bipolar transistor and the comparable Siliconix VNE003A *n*-channel MOSFET. For $V_{CE} > 10$ volts, second breakdown limits the *npn* transistor dc collector current to values corresponding to less than the maximum allowable dissipation of 250 watts. The problem is less severe for short pulses, and it effectively disappears for pulses of 1ms duration or less. Note that the MOSFET has no second breakdown; its SOA is bounded by maximum current (bonding-wire limited, therefore higher for short pulses), maximum dissipation, and maximum allowable drain-source voltage. See Chapter 3 for more details on power MOSFETs.

□ 6.08 Programmable supplies

There is frequently the need for power supplies that can be adjusted right down to zero volts, especially in bench applications where a flexible source of power is essential. In addition, it is often desirable to be able to “program” the output voltage with another voltage or with a digital input (via digital thumbwheel switches, for instance).

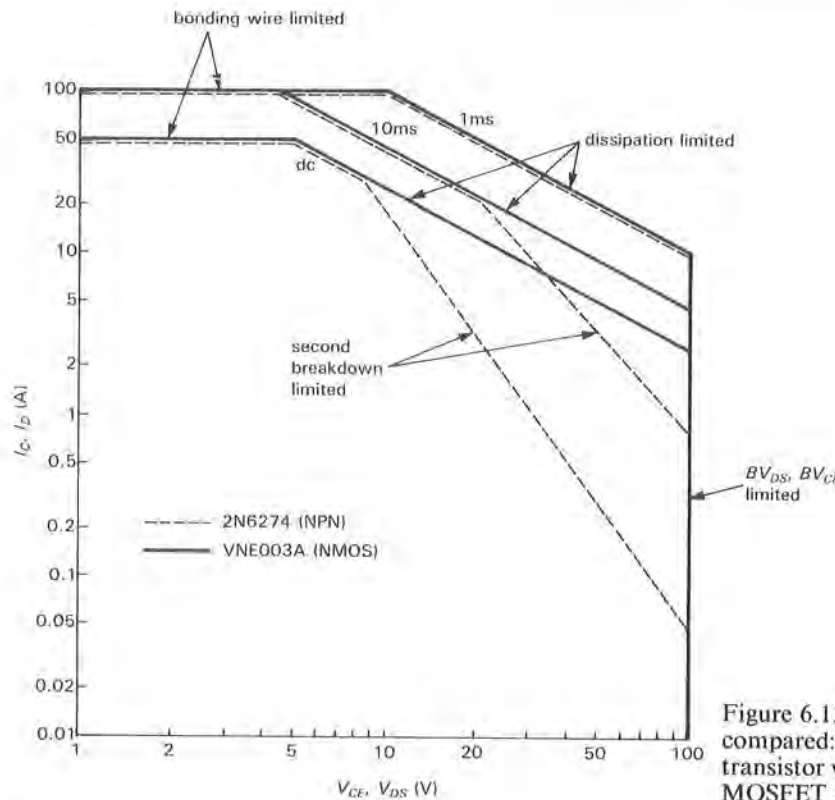


Figure 6.13. Safe operating areas compared: bipolar *nnp* power transistor versus *n*-channel power MOSFET of same ratings.

Figure 6.14 shows the classic scheme for a supply that is adjustable down to zero output voltage (as our 723 circuits so far are not). A separate split supply provides

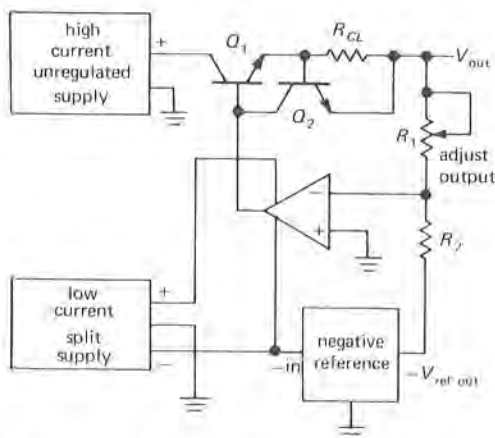


Figure 6.14. Regulator adjustable down to zero volts.

power for the regulator and also generates an accurate negative reference voltage (more on references in Sections 6.14 and 6.15). R_1 sets the output voltage (since the inverting input will be at ground), which can be adjusted all the way down to zero (at zero resistance). When the regulator circuitry (which can be an integrated circuit or discrete components) is run from a split supply, no problems are encountered at low output voltages.

To make the supply programmable with an external voltage, just replace V_{ref} with an externally controlled voltage (Fig. 6.15). The rest of the circuit is unchanged. R_1 now sets the scale of $V_{control}$.

Digital programmability can be added by replacing V_{ref} with a device called a DAC (digital-to-analog converter) with current-sinking output. These devices, which we will discuss later, convert a binary input code to a proportional current

(or voltage) output. A good choice here is the AD7548, a monolithic 12-bit DAC with current-sinking output and a price tag of about \$9. By replacing R_2 with the DAC, you get a digitally programmed supply, with step size of $1/4096$ (2^{-12}) of the full-scale output voltage. Since the inverting input is a virtual ground, the DAC doesn't even have to have any output compliance. In practice, R_1 would be adjusted to set a convenient scale for the output, say 1mV per input digit.

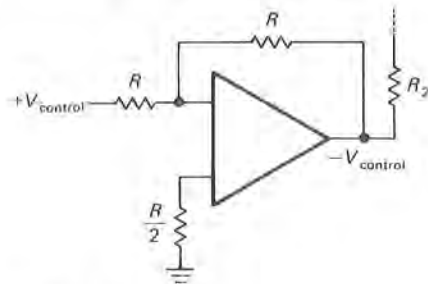


Figure 6.15

□ 6.09 Power-supply circuit example

The “laboratory” bench supply shown in Figure 6.16 should help pull all these design ideas together. It is important to be able to adjust the regulated output voltage right down to zero volts in a general-purpose bench supply, so an additional split supply is used to power the regulator. IC_1 is a high-voltage op-amp, which can operate with 80 volts total supply voltage. We used paralleled power MOSFETs as the output pass transistor, both because of its easy gate drive requirements and its excellent safe operating area (characteristic of all power MOSFETs). The combination can dissipate plenty of power (60W per transistor at 100°C case temperature), which is necessary even for moderate output current when such a wide range of output voltage is provided. This is because the unregulated input voltage has to be high enough for the maximum regulated

output voltage, resulting in a large voltage drop across the pass transistors when the regulated output voltage is low. Some supplies solve this problem by having several ranges of output voltage, switching the unregulated input voltage accordingly. There are even supplies with the unregulated supply driven from a variable-voltage transformer ganged to the same control as the output voltage. In both cases you lose the capability of remote programmability.

EXERCISE 6.4

What is the maximum power dissipation in the pass transistors for this circuit?

R_1 is a precision multidecade potentiometer for precise and linear adjustment of the output voltage. The output voltage is referenced to the 1N829 precision zener (5ppm/°C tempco at 7.5mA zener current). The current-limiting circuitry is considerably better than the simple protective current limiters we have been discussing, since it is sometimes desirable to be able to set a precise and stable current limit when using a bench supply. Note the unusual (but convenient) method of current limiting by sinking current from the compensation pin of IC_1 , which has unity gain to the output while operating at low current. By providing both precision-regulated voltage (all the way down to 0V) and current, the device becomes a flexible laboratory power supply. With this current-limit method, the supply becomes a flexible constant-current source. Q_4 provides a constant 100mA load, maintaining good performance near zero output voltage (or current) by keeping the pass transistors well into the active region. This current sink also allows the load to source some current into the supply without its output voltage rising. This is useful with the bizarre loads you sometimes encounter, e.g., an instrument that contains some additional supplies of its own capable of sourcing some current into the power-supply output terminal.

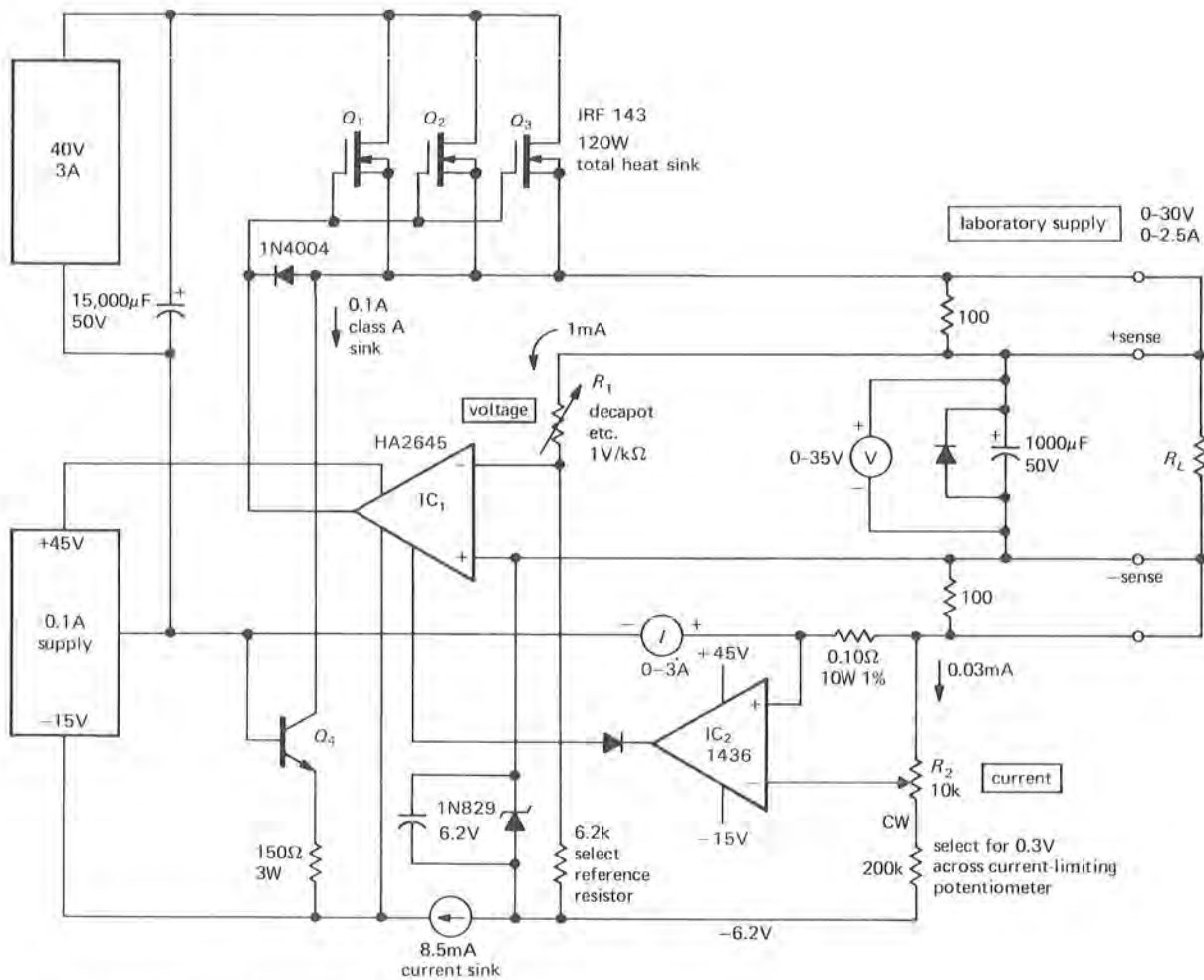


Figure 6.16. Laboratory bench supply.

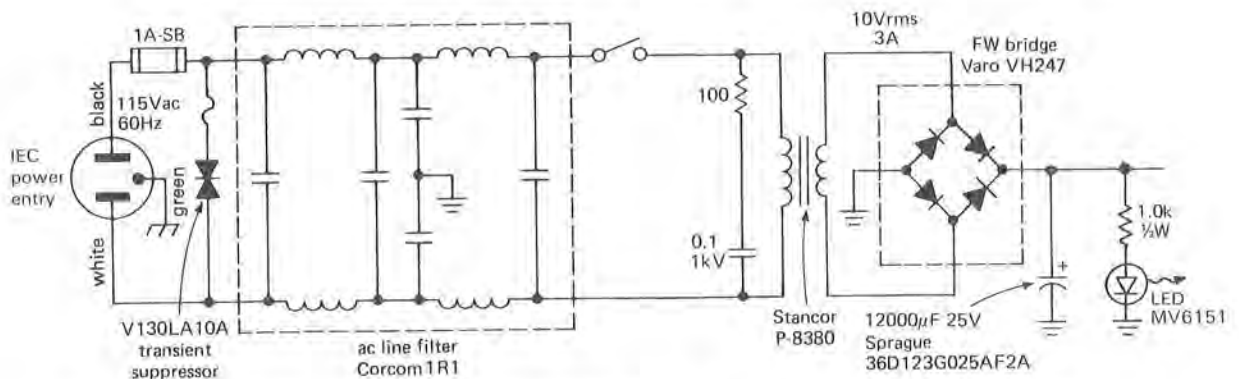


Figure 6.17. Unregulated supply with ac line connections. Note color convention of ac line cord.

Note the external sense leads, with default connection to the power-supply output terminals. For precise regulation of output voltage at the load, you would bring external sense leads to the load itself, eliminating (through feedback) voltage drops in the connecting leads.

6.10 Other regulator ICs

The 723 was the original voltage regulator IC, and it is still a useful chip. There are a few improved versions that work much the same way, however, and you should consider them when you design a regulated power supply. The LAS1000 and LAS1100 from Lambda and the SG3532 from Silicon General can operate down to 4.5 volts input voltage, because they use an internal 2.5 volt “bandgap reference” (see Section 6.15) rather than the 7.15 volt zener of the 723. They also have internal circuitry that shuts off the chip if it overheats; compare the 723’s solution (burnout!). Although these regulators have the same pin names, you can’t just plug these regulators into a socket intended for a 723, because (among other things) they assume a lower reference voltage. Another 723-like regulator is the MC1469 (and its negative twin, the MC1463) from Motorola.

If you look at modern power-supply circuits, you won’t see many 723s, or even the improved versions we just mentioned. Instead, you’ll see mostly ICs like the 7805 or 317, with a remarkable absence of external components (the 7805 requires none!). Most of the time you can get all the performance you need from these highly integrated and easy-to-use “three-terminal” regulators, including high output current (up to 10A) without external pass transistors, adjustable output voltage, excellent regulation, and internal current limiting and thermal shut-down. We’ll talk about these shortly, but first an interlude on (a) the design of the unregulated supply and (b) voltage references.

THE UNREGULATED SUPPLY

All regulated supplies require a source of “unregulated” dc, a subject we introduced in Section 1.27 in connection with rectifiers and ripple calculations. Let’s look at this subject in more detail, beginning with the circuit shown in Figure 6.17. This is an unregulated +13 volt (nominal) supply for use with a +5 volt 2 amp regulator. Let’s go through it from left to right, pointing out some of the things to keep in mind when you do this sort of design.

6.11 ac line components

Three-wire connection

Always use a 3-wire line cord with neutral (green) connected to the instrument case. Instruments with ungrounded cases can become lethal devices in the event of transformer insulation failure or accidental connection of one side of the power line to the case. With a grounded case, such a failure simply blows a fuse. You often see instruments with the line cord attached to the chassis (permanently) using a plastic "strain relief," made by Heyco or Richco. A better way is to use an IEC three-prong male chassis-mounted connector, to mate with those popular line cords that have the three-prong IEC female molded onto the end. That way the line cord is conveniently removable. Better yet, you can get a combined "power entry module," containing IEC connector, fuse holder, line filter, and switch (as described later). Note that ac wiring uses a nonintuitive color convention: black = "hot," white = neutral, and green = ground.

Line filter and transient suppressor

In this supply we have used a simple *LC* line filter. Although they are often omitted, such filters are a good idea, since they serve the purpose of preventing possible radiation of radiofrequency interference

(RFI) from the instrument via the power line, as well as filtering out incoming interference that may be present on the power line. Power-line filters with excellent performance characteristics are available from several manufacturers, e.g., Corcom, Cornell-Dubilier, and Sprague. Studies have shown that spikes as large as 1kV to 5kV are occasionally present on the power lines at most locations, and smaller spikes occur quite frequently. Line filters are reasonably effective in reducing such interference.

In many situations it is desirable to use a "transient suppressor," as shown, a device that conducts when its terminal voltage exceeds certain limits (it's like a bidirectional high-power zener). These are inexpensive and small and can short out hundreds of amperes of potentially harmful current in the form of spikes. Transient suppressors are made by a number of companies, e.g., GE and Siemens. Tables 6.2 and 6.3 list some useful RFI filters and transient suppressors.

Fuse

A fuse is essential in every piece of electronic equipment. The large wall fuses or circuit breakers (typically 15–20A) in house or lab won't protect electronic equipment, since they are chosen to blow only when the current rating of the wiring in the wall is exceeded. For instance, a

TABLE 6.2. 130 VOLT AC TRANSIENT SUPPRESSORS

Type	Manuf.	Diameter (in)	Energy (W-s)	Peak curr (A)	Capacitance (pF)
V130LA1	GE	0.34	4	500	180
S07K130	Siemens	0.35	6	500	130
V130LA10A	GE	0.65	30	4000	1000
S14K130	Siemens	0.67	22	2000	1000
V130LA20B	GE	0.89	50	6000	1900
S20K130	Siemens	0.91	44	4000	2300

TABLE 6.3. 115 VOLT AC POWER FILTERS (IEC CONNECTOR^a)

Manuf.	Part No.	Circuit	Current (A)	Attenuation ^b (line-to-gnd, 50Ω/50Ω)			Comments
				150kHz (dB)	500kHz (dB)	1MHz (dB)	
Corcom	3EF1	π	3	15	25	30	general purpose
	3EC1	π	3	20	30	37	higher attenuation
	3EDSC2-2	π	3	32	37	44	with fuse
	2EDL1S	π	2	14	—	24	with fuse and switch
Curtis	F2100CA03	π	3	15	25	30	general purpose
	F2400CA03	π	3	22	35	40	higher attenuation
	F2600FA03	π	3	21	35	41	with fuse
	PE810103	π	3	18	24	30	with fuse and switch
Delta	03GEEG3H	π	3	24	30	38	general purpose
	03SEEG3H	dual-π	3	42	65	70	higher attenuation
	04BEEG3H	π	4	26	35	40	with fuse
	03CK2	π	3	35	40	40	with fuse and switch
	03CR2	dual-π	3	50	60	55	same, higher attenuation
Schaffner	FN323-3	π	3	22	32	36	general purpose
	FN321-3	π	3	35	43	46	higher attenuation
	FN361-2	π	2	25	40	46	with fuse
	FN291-2.5	π	2.5	25	40	46	with fuse and switch
	FN1393-2.5	π	2.5	40	45	42	same, higher attenuation
Sprague	3JX5421A	π	3	15	25	30	general purpose
	3JX5425C	π	3	20	30	37	higher attenuation
	200JM6-2	π	6	12	25	—	with fuse

(a) these units are representative of a large selection, many of which do not include an IEC input connector. (b) rf attenuation figures are measured in a 50Ω system, and should not be relied upon to predict performance in an ac line circuit.

house wired with 14 gauge wire will have 15 amp breakers. Now, if the filter capacitor in the preceding supply becomes short-circuited someday (a typical failure mode), the transformer might then draw 5 amps primary current (instead of its usual 0.25A). The house breaker won't open, but your instrument becomes an incendiary device, with its transformer dissipating over 500 watts!

Some notes on fuses: (a) It is best to use a "slow-blow" type in the power-line circuit, because there is invariably a large current transient at turn-on (caused mostly by rapid charging of the power-supply filter capacitors). (b) You may think you know how to calculate the fuse current rating,

but you're probably wrong. A dc power supply has a high ratio of rms current to average current, because of the small conduction angle (fraction of the cycle over which the diodes are conducting). The problem is worse if overly large filter capacitors are used. The result is an rms current considerably higher than you would estimate. The best procedure is to use a "true rms" ac current meter to measure the actual rms line current, then choose a fuse of at least 50% higher current rating (to allow for high line voltage, the effects of fuse "fatigue," etc.). (c) When wiring cartridge-type fuse holders (used with the popular 3AG fuse, which is almost universal in electronic equipment), be sure

to connect the leads so that anyone changing the fuse cannot come in contact with the power line. This means connecting the “hot” lead to the rear terminal of the fuse holder (the authors learned this the hard way!). Commercial power-entry modules with integral fuse holders are usually arranged so that the fuse cannot be reached without removing the line cord.

Shock hazard

Incidentally, it is a good idea to insulate all exposed 110 volt power connections inside any instrument, using Teflon heat-shrink tubing, for instance (the use of “friction tape” or electrical tape inside electronic instruments is strictly bush-league). Since most transistorized circuits operate on relatively low dc voltages ($\pm 15\text{V}$ to $\pm 30\text{V}$ or so), from which it is not possible to receive a shock, the power line wiring is the only place where any shock hazard exists in most electronic devices (there are exceptions, of course). The front-panel ON/OFF switch is particularly insidious in this respect, since it is close to other low-voltage wiring. Your test instruments (or, worse, your fingers) can easily come in contact with it when you go to pick up the instrument while testing it.

Miscellany

We favor “power-entry modules,” combining a 3-prong IEC connector (use a removable line cord) and some combination of line filter, fuse holder, and power switch. For example, the Schaffner FN380 series (or Corcom L series) has all these features, and they are available with maximum currents from 2 to 6 amps. They give you options for fusing or switching either one or both sides of the line, and they offer several filter configurations. Some other manufacturers offering similar products are Curtis, Delta, and Power Dynamics (Table 6.3).

Our circuit shows an LED pilot light (with current-limiting resistor) running from the unregulated dc voltage. It is generally better practice to power the LED from the regulated dc, so that it doesn’t flicker with load or power-line variations.

The series combination of 100 ohms and $0.1\mu\text{F}$ capacitor across the transformer primary prevents the large inductive transient that would otherwise occur at turn-off. This is often omitted, but it is highly desirable, particularly in equipment intended for use near computers or other digital devices. Sometimes this *RC* “snubber” network is wired across the switch, which is equivalent.

6.12 Transformer

Now for the transformer. Never build an instrument to run off the power line without a transformer! To do so is to flirt with disaster. Transformerless power supplies, which are popular in some consumer electronics (radios and televisions, particularly) because they’re cheap, put the circuit at high voltage with respect to external ground (water pipes, etc.). This has no place in instruments intended to interconnect with any other equipment and should always be avoided. And use extreme caution when servicing any such equipment; just connecting your oscilloscope probe to the chassis can be a shocking experience.

The choice of transformer is more involved than you might at first expect. One problem is that manufacturers have been slow to introduce transformers with voltages and currents appropriate for transistorized circuitry (the catalogs are still cluttered with transformers designed for vacuum tubes), and you wind up making compromises you’d rather avoid. We have found the Signal Transformer Company unusual, with their nice selection of transformers and quick delivery. Don’t overlook the possibility of having

transformers custom-made if your application requires more than a few.

Even assuming that you can get the transformer you want, you still have to decide what voltage and current are best. The lower the input voltage to the regulator, the lower the dissipation in the pass transistors. But you must be absolutely certain the input to the regulator will never drop below the minimum necessary for regulation, typically 2 to 3 volts above the regulated output voltage, or you may encounter 120Hz dips in the regulated output. The amount of ripple in the unregulated output is involved here, since it is the minimum input to the regulator that must stay above some critical voltage, but it is the average input to the regulator that determines the transistor dissipation.

As an example, for a +5 volt regulator you might use an unregulated input of +10 volts at the minimum of the ripple, which itself might be a volt or two. From the secondary voltage rating you can make a pretty good guess of the dc output from the bridge, since the peak voltage (at the top of the ripple) is approximately 1.4 times the rms secondary voltage, less two diode drops. But it is essential to make actual measurements if you are designing a power supply with near-minimum drop across the regulator, because the actual output voltage of the unregulated supply depends on poorly specified parameters of the transformer, such as winding resistance and magnetic coupling, both of which contribute to voltage drop under load. Be sure to make measurements under worst-case conditions: full load and low power-line voltage (105V). Remember that large filter capacitors typically have loose tolerances: -30% to +100% about the nominal value is not unusual. It is a good idea to use transformers with multiple taps on the primary, when available, for final adjustment of output voltage. The Triad F-90X series and the Stancor TP series are very flexible this way.

One further note on transformers: Current ratings are sometimes given as rms secondary current, particularly for transformers intended for use into a resistive load (filament transformers, for instance). Since a rectifier circuit draws current only over a small part of the cycle (during the time the capacitor is actually charging), the rms current, and therefore the I^2R heating, is likely to exceed specifications for a load current approaching the rated rms current of the transformer. The situation gets worse as you increase capacitor size to reduce preregulator ripple; this simply requires a transformer of larger rating. Full-wave rectification is better in this respect, since a greater portion of the transformer waveform is used.

6.13 dc components

Filter capacitor

The filter capacitor is chosen large enough to provide acceptably low ripple voltage, with voltage rating sufficient to handle the worst-case combination of no load and high line voltage (125–130V rms). For the circuit shown in Figure 6.17, the ripple is about 1.5 volts pp at full load. Good design practice calls for the use of computer-type electrolytics (they come in a cylindrical package with screw terminals at one end), e.g., the Sprague 36D type. In smaller capacitance values most manufacturers provide capacitors of equivalent quality in an axial-lead package (one wire sticking out each end), e.g., the Sprague 39D type. Watch out for the loose capacitance tolerance!

At this point it may be helpful to look back at Section 1.27, where we first discussed the subject of ripple. With the exception of switching regulators (see Section 6.19 and following), you can always calculate ripple voltage by assuming a constant-current load equal to the maximum output load current. In fact, the input to a series

regulator looks just like a constant-current sink. This simplifies your arithmetic, since the capacitor discharges with a ramp, and you don't have to worry about time constants or exponentials (Fig. 6.18).

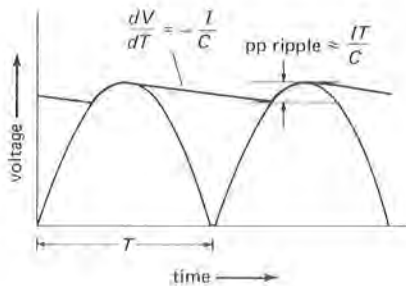


Figure 6.18

For example, suppose you want to choose a filter capacitor for the unregulated portion of a +5 volt 1 amp regulated supply, and suppose you have already chosen a transformer with a 10 volt rms secondary, to give an unregulated dc output of 12 volts (at the peak of the ripple) at full load current. With a typical regulator dropout voltage of 2 volts, the input to the regulator should never dip below +7 volts (the 723 will require +9.5V, but the convenient 3-terminal regulators discussed in Section 6.16 are more friendly). Since you have to contend with a $\pm 10\%$ worst-case line-voltage variation, you should keep ripple to less than 2 volts pp. Therefore,

$$2 = T(dV/dT) = TI/C = 0.008 \times 1.0/C$$

from which $C = 4000\mu\text{F}$. A $5000\mu\text{F}$ 25 volt electrolytic would be a minimum choice, with allowance for a 20% tolerance in capacitor value. When choosing filter capacitors, don't get carried away: An oversize capacitor not only wastes space but also increases transformer heating (by reducing the conduction angle, hence increasing the ratio $I_{\text{rms}}/I_{\text{avg}}$). It also increases stress on the rectifiers.

The LED shown across the output in Figure 6.17 acts as a "bleeder" to discharge

the capacitor in a few seconds under no-load conditions. This is a good feature, because power supplies that stay charged after things have been shut off can easily lead you to damage some circuit components if you mistakenly think that no voltage is present.

Rectifier

The first point to be made is that the diodes used in power supplies are quite different from the small 1N914-type signal diodes used in circuitry. Signal diodes are generally designed for high speed (a few nanoseconds), low leakage (a few nanoamps), and low capacitance (a few picofarads), and they can generally handle currents up to about 100mA, with breakdown voltages rarely exceeding 100 volts. By contrast, rectifier diodes and bridges for use in power supplies are hefty objects with current ratings going from 1 amp to 25 amps or more and breakdown voltages going from 100 volts to 1000 volts. They have relatively high leakage currents (in the range of microamps to milliamps) and plenty of junction capacitance. They are not intended for high speed. Table 6.4 lists a selection of popular types.

Typical of rectifiers is the popular 1N4001–1N4007 series, rated at 1 amp, with reverse-breakdown voltages ranging from 50 to 1000 volts. The 1N5625 series is rated at 3 amps, which is about the highest current available in a lead-mounted (cooled by conduction through the leads) package. The popular 1N1183A series typifies high-current stud-mounted rectifiers, with a current rating of 40 amps and breakdown voltages to 600 volts. Plastic-encapsulated bridge rectifiers are quite popular also, with lead-mounted 1 and 2 amp types and chassis-mounted packages in ratings up to 25 amps or more. For rectifier applications where high speed is important (e.g., dc-to-dc converters, see Section 6.19), fast-recovery diodes are

TABLE 6.4. RECTIFIERS

Type	Breakdown voltage V_{BR} (V)	Forward drop V_F typ (V)	@	Average current I_o (A)	Package	Comments
General purpose						
1N4001-07	50-1000	0.9		1	lead-mounted	popular
1N5059-62	200-800	1.0		2	lead-mounted	
1N5624-27	200-800	1.0		5	lead-mounted	
1N1183A-90A	50-600	0.9		40	stud-mounted	popular; -R for rev. pol.
Fast recovery ($t_r = 0.1\mu s$ typ)						
1N4933-37	50-600	1.0		1	lead-mounted	
1N5415-19	50-500	1.0		3	lead-mounted	
1N3879-83	50-400	1.2		6	stud-mounted	-R for reverse polarity
1N5832-34	50-400	1.0		20	stud-mounted	-R for reverse polarity
Schottky (low V_F , very fast)						
1N5817-19	20-40	0.6^m		1	lead-mounted	
1N5820-22	20-40	0.5^m		3	lead-mounted	
1N5826-28	20-40	0.5^m		15	stud-mounted	
1N5832-34	20-40	0.6^m		40	stud-mounted	
Full-wave bridge						
3N246-52	50-1000	0.9		1	plastic SIP	MDA100A
3N253-59	50-1000			2	plastic SIP	MDA200
MDA970A1-A5	50-400	0.85		8	chassis mtd	
MDA3500-10	50-1000			35	chassis mtd	
Exotic						
GE A570A-A640L	100-2000	1.0^m		1500	giant button	high current!
Semtech SCH5000-25000	5kV-25kV	$7-33^m$		0.5	lead-mounted	HV, curr; fast ($0.2\mu s$)
Varo VF25-5 to -40	5kV-40kV	$12-50^m$		0.025	lead-mounted	high voltage
Semtech SCKV100K3-200K3	100kV-200kV	$150-300$		0.1	plastic rod	very high voltage

(m) maximum.

available, e.g., the 1N4933 series of 1 amp diodes. For low-voltage applications it may be desirable to use Schottky barrier rectifiers, e.g., the 1N5823 series, with forward drops of less than 0.4 volt at 5 amps.

VOLTAGE REFERENCES

There is frequently the need for good voltage references within a circuit. For instance, you might wish to construct a precision regulated supply with characteristics

better than those you can obtain using complete regulators like the 723 (since integrated voltage regulator chips usually dissipate considerable power because of the built-in pass transistor, they tend to heat up, with consequent drift). Or you might want to construct a precision constant-current supply. Another application that requires a precision reference, but not a precision power supply, is design of an accurate voltmeter, ohmmeter, or ammeter.

There are two kinds of voltage references — *zener diodes* and *bandgap*

references; each can be used alone or as an internal part of an integrated circuit voltage reference.

□ 6.14 Zener diodes

The simplest form of voltage reference is the zener diode, a device we discussed in Section 1.06. Basically, it is a diode operated in the reverse-bias region, where current begins to flow at some voltage and increases dramatically with further increases in voltage. To use it as a reference, you simply provide a roughly constant current; this is often done with a resistor from a higher supply voltage, forming the most primitive kind of regulated supply.

Zeners are available in selected voltages from 2 to 200 volts (they come in the same series of values as standard 5% resistors), with power ratings from a fraction of a watt to 50 watts and tolerances of 1% to 20%. As attractive as they might seem for use as general-purpose voltage references, zeners are actually somewhat difficult to use, for a variety of reasons. It is necessary to stock a selection of values, the voltage tolerance is poor except in high-priced precision zeners, they are noisy, and the zener voltage depends on current and temperature. As an example of the last two effects, a 27 volt zener in the popular 1N5221 series of 500mW zeners has a temperature coefficient of $+0.1\%/^{\circ}\text{C}$, and it will change voltage by 1% when its current varies from 10% to 50% of maximum.

There is an exception to this generally poor performance of zeners. It turns out that in the neighborhood of 6 volts, zener diodes become very stiff against changes in current and simultaneously achieve a nearly zero temperature coefficient. The graphs in Figure 6.19, plotted from measurements on zeners with different voltages, illustrate the effects. This peculiar behavior comes about because "zener" diodes actually employ two different

mechanisms: zener breakdown (low voltage) and avalanche breakdown (high voltage). If you need a zener for use as a stable voltage reference only, and you don't care what voltage it is, the best thing to use is one of the compensated zener references constructed from a 5.6 volts zener (approximately) in series with a forward-biased diode. The zener voltage is chosen to give a positive coefficient to cancel the diode's temperature coefficient of $-2.1\text{mV}/^{\circ}\text{C}$.

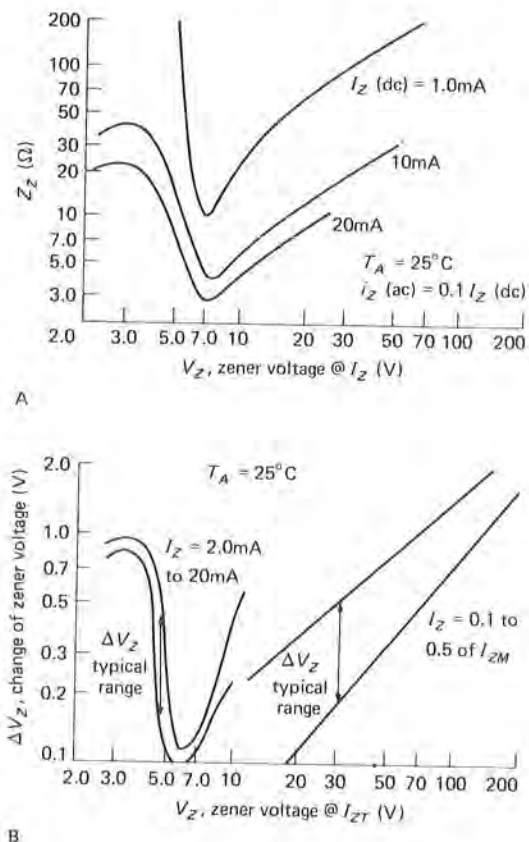


Figure 6.19. Zener diode impedance and regulation for zener diodes of various voltages. (Courtesy of Motorola, Inc.)

As you can see from the graph in Figure 6.20, the temperature coefficient depends on operating current and also on the zener voltage. Therefore, by choosing the

zener current properly, you can “tune” the temperature coefficient somewhat. Such zeners with built-in series diodes make particularly good references. As an example, the 1N821 series of inexpensive 6.2 volt references offers temperature coefficients going from 100ppm/°C (1N821) down to 5ppm/°C (1N829); the 1N940 and 1N946 are 9 volt and 11.7 volt references with tempcos of 2ppm/°C.

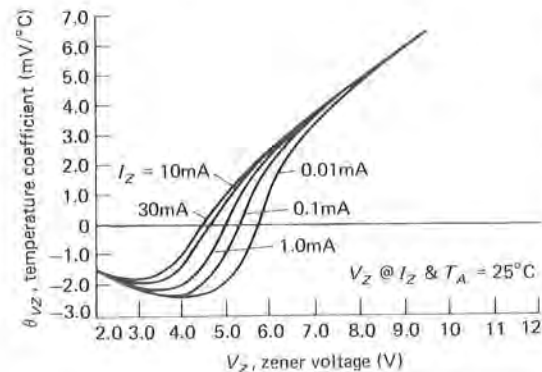


Figure 6.20. Temperature coefficient of zener diode breakdown voltage versus the voltage of the zener diode. (Courtesy of Motorola, Inc.)

□ Providing operating current

These compensated zeners can be used as stable voltage references within a circuit, but they must be provided with constant current. The 1N821 series is specified as 6.2 volts $\pm 5\%$ at 7.5mA, with an incremental resistance of about 15 ohms; thus, a change in current of 1mA changes the reference voltage three times as much as a change in temperature from -55°C to $+100^{\circ}\text{C}$ for the 1N829. Figure 6.21 shows a simple way to provide constant bias current for a precision zener. The op-amp is wired as a noninverting amplifier in order to generate an output of exactly +10.0 volts. That stable output is itself used to provide a precision 7.5mA bias current. This circuit is self-starting, but it can turn on with either polarity of

output! For the “wrong” polarity, the zener operates as an ordinary forward-biased diode. Running the op-amp from a single supply, as shown, overcomes this bizarre problem. Be sure to use an op-amp that has common-mode input range to the negative rail (“single-supply” op-amps).

There are special compensated zeners available with guaranteed stability of zener voltage with *time*, a specification that normally tends to get left out. Examples are the 1N3501 and 1N4890 series. Zeners of this type are available with guaranteed stability of better than 5ppm/1000h. They’re not cheap. Table 6.5 lists the characteristics of some useful zeners and reference diodes, and Table 6.6 shows part numbers for two popular 500mW general-purpose zener families.

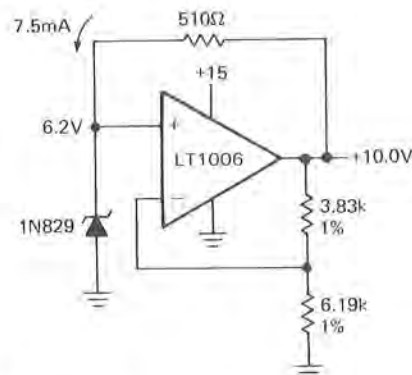


Figure 6.21

IC zeners

The 723 regulator uses a compensated zener reference to achieve its excellent performance (30ppm/°C stability of V_{ref}). The 723, in fact, is quite respectable as a voltage reference all by itself, and you can use the other components of the IC to generate a stable reference output at any desired voltage.

The 723 used as a voltage reference is an example of a *3-terminal reference*, meaning that it requires a power supply to

TABLE 6.5. ZENER AND REFERENCE DIODES^a

Type	Zener voltage V_Z (V)	Test current @ I_{ZT} (mA)	Tolerance ($\pm\%$)	Tempco max (ppm/ $^{\circ}$ C)	Regulation ΔV for $\pm 10\% I_{ZT}$ max (mV)	P_{diss} max (W)	Comments
Reference zeners							
1N821A-	6.2	7.5	5	± 100	7.5	0.4	5 member family, graded by tempco; best and worst shown
1N829A	6.2	7.5	5	± 5	7.5	0.4	
1N4890-	6.35	7.5	5	± 20		0.4	long-term stab < 100ppm/1000h
1N4895	6.35	7.5	5	± 5		0.4	long-term stab < 10ppm/1000h
Regulator zeners							
1N5221A	2.4	20	10	-850	60	0.5	60 member family, 2.4V to 200V, in "5% resistor values," plus some extras. -B = $\pm 5\%$; popular ^b
1N5231A	5.1	20	10	± 300	34	0.5	
1N5281A	200	0.65	10	+1100	160	0.5	
1N4728A	3.3	76	10	-750	76	1.0	37 member family, 3.3V to 100V, in "5% resistor values."
1N4735A	6.2	41	10	+500	8	1.0	
1N4764A	100	2.5	10	+1100	88	1.0	

(^a) see also Table 6.7 (IC Voltage References). (^b) see Table 6.6 (500mW Zeners).

operate, and includes internal circuitry to bias the zener and buffer the output voltage. Improved 3-terminal IC zeners include the excellent LM369 from National (1.5ppm/ $^{\circ}$ C typ), and the REF10KM from Burr-Brown (1ppm/ $^{\circ}$ C max tempco); we've often used the inexpensive Motorola MC1404 (which is actually a bandgap reference, see below) in our circuits. We'll treat 3-terminal precision references in more detail shortly, after discussing the simpler 2-terminal types.

Precision temperature-compensated zener ICs are available as *2-terminal references* also; electrically they look just like zeners, although they actually include a number of active devices to give improved performance (most notably, constancy of "zener" voltage with applied current). An example is the inexpensive LM329, with a zener voltage of 6.9 volts. Its best version has a temperature coefficient of 6ppm/ $^{\circ}$ C (typ), 10ppm/ $^{\circ}$ C (max), when provided with a constant current of 1mA.

Some unusual IC zeners include the temperature-stabilized LM399 (0.3ppm/ $^{\circ}$ C

TABLE 6.6. 500mW ZENER DIODES

1N5221 series	1N746 series	V_Z (V)	@ I_{ZT} (mA)
1N5230	1N750	4.7	20
1N5231	1N751	5.1	20
1N5232	1N752	5.6	20
1N5233	—	6.0	20
1N5235	1N754	6.8	20
1N5236	1N755	7.5	20
1N5237	1N756	8.2	20
1N5240	1N758	10	20
1N5242	1N759	12	20
1N5245	1N965	15	8.5
1N5248	1N967	18	7.0
1N5250	1N968	20	6.2
1N5253	—	25	5.0
1N5256	1N972	30	4.2
1N5259	1N975	39	3.2
1N5261	1N977	47	2.7
1N5267	1N982	75	1.7
1N5271	1N985	100	1.3
1N5276	1N989	150	0.85
1N5281	1N992	200	0.65

typ), the micropower LM385 (which operates down to $10\mu\text{A}$), and the astounding LTZ1000 from Linear Technology, with its $0.05\text{ppm}/^\circ\text{C}$ typical tempco, 0.3ppm per square-root-month drift, and $1.2\mu\text{V}$ low-frequency noise.

Zener diodes can be very noisy, and some IC zeners suffer from the same disease. The noise is related to surface effects, however, and *buried* (or *subsurface*) zener diodes are considerably quieter. In fact, the LTZ1000 buried zener just mentioned is the quietest reference of any kind. The LM369 and REF10KM also have very low noise.

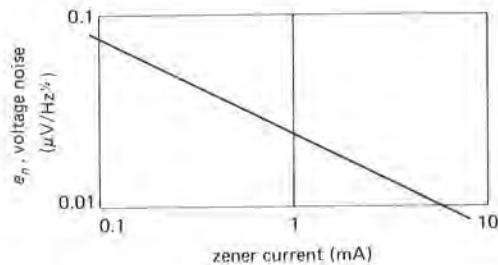


Figure 6.22. Voltage noise for a low-noise zener reference diode similar to the type used in the 723 regulator.

Table 6.7 lists the characteristics of nearly all available IC references, both zener and bandgap.

□ 6.15 Bandgap (V_{BE}) reference

More recently, a circuit known as a “band-gap” reference has become popular. It should properly be called a V_{BE} reference, and it is easily understandable using the Ebers-Moll diode equation. Basically, it involves the generation of a voltage with a positive temperature coefficient the same as V_{BE} ’s negative coefficient; when added to a V_{BE} , the resultant voltage has zero tempco.

We start with a current mirror with two transistors operating at different emitter

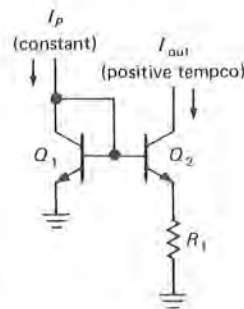


Figure 6.23

current densities (typically a ratio of 10:1) (see Fig. 6.23). Using the Ebers-Moll equation, it is easy to show that I_{out} has a positive temperature coefficient, since the difference in V_{BE} ’s is just $(kT/q) \log_e r$, where r is the ratio of current densities (see the graph in Fig. 2.53). You may wonder where we get the constant programming current I_P . Don’t worry; you’ll see the clever method at the end. Now all you do is convert that current to a voltage with a resistor and add a normal V_{BE} .

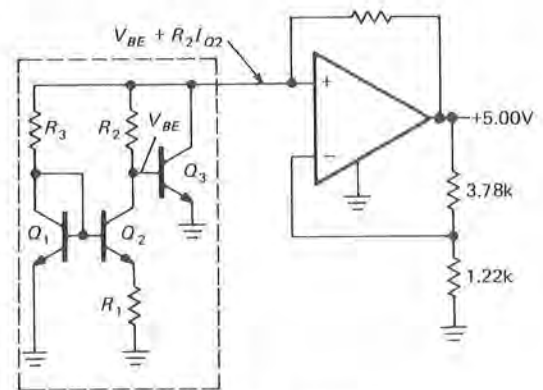


Figure 6.24. Classic V_{BE} bandgap voltage reference.

Figure 6.24 shows the circuit. R_2 sets the amount of positive-coefficient voltage you have added to V_{BE} , and by choosing it appropriately, you get zero overall temperature coefficient. It turns out that zero temperature coefficient occurs when the

TABLE 6.7. IC VOLTAGE REFERENCES

Type	Mfg ^a	B ⁺ gap/zener	Terminals	Trim	Voltage (V)	Acc'y (%)	Tempco typ (ppm/°C)	Min supply voltage (V)	Supply curr (mA)	Output curr max (mA)	Noise voltage 0.1-10Hz typ (μV pp)	Long-term stability typ (ppm/1000h)	Regulation	
													Line typ (%/V)	Load 0-10mA typ (%)
Regulator type														
LM10C	NS+	B	8	•	0.20	5	30	1.1	0.3	20	—	—	0.001	0.01 ^a
μA723C	FS+	Z	14	•	7.15	3	20	9.5	2.3	65	—	1000	0.003	0.03
SG3532J	SG+	B	10	•	2.50	4	50	4.5	1.6	150	—	300	0.005	0.02
Two-terminal (zener) type														
LM129A	NS	Z	2	—	6.9	5	6	—	1	15 ^b	—	20	—	0.1
VR182C	DA	B	2	—	2.455	1.4	23	—	2	120 ^b	10 ^e	10	—	0.1
LM313	NS	B	2	—	1.22	5	100	—	1 ^f	20 ^b	5 ^f	—	—	0.5 ^f
LM329C	NS	Z	2	—	6.9	5	30	—	1	15 ^b	—	20	—	0.1
LM336-2.5	NS	B	3	•	2.50	4	10	—	1	10 ^b	—	20	—	0.1
LM336B-5	NS	B	3	•	5.0	1	15	—	1 ^o	10 ^b	—	20	—	0.1 ^o
LM385B	NS	B	2	—	1.23	1	20	—	0.1 ^a	20 ^b	25	—	—	0.02 ^a
LM385BX-1.2	NS	B	2	—	1.235	1	30 ^m	—	0.1 ^d	20 ^b	60 ^f	20	—	0.8 ⁿ
LM385BX-2.5	NS	B	2	—	2.50	1.5	30 ^m	—	0.1 ^p	20 ^b	120 ^f	20	—	0.4 ⁿ
LM299A	NS	Z	4	—	6.95	2	0.2	9	17	10 ^b	—	20	—	0.1
LM399	NS	Z	4	—	6.95	5	0.3	9	17	10 ^b	—	20	—	0.1
LM3999	NS	Z	3	—	6.95	5	2.0	9	17	10 ^b	—	20	—	0.1
TL430	TI	B	3	•	2.75	5	120	—	10	100 ^b	50	—	—	0.5
TL431	TI	B	3	•	2.75	2	10	—	10	100 ^b	50	—	—	0.5
AD589M	AD	B	2	—	1.235	2	10 ^m	—	0.1 ^h	5 ^b	5 ^f	—	—	0.05 ^a
LTZ1000	LT	Z	2	—	7.2	4	0.05	—	5	—	1.2	0.3 ^s	—	1 ^f
LT1004C-1.2	LT	B	2	—	1.235	0.3	20	—	0.1 ^d	20 ^b	60 ^f	20	—	0.8 ^{mn}
LT1009C	LT	B	3	•	2.50	0.2	15	—	1 ^o	10 ^b	—	20	—	0.1 ^o
LT1029A	LT	B	3	•	5.0	0.2	8	—	1 ^o	10 ^b	—	20	—	0.04 ^o
LT1034B	LT	B	3	—	1.225	1	10	—	0.1 ^p	20 ^b	4	—	—	0.3 ^p
"	LT	Z	—	—	7.0	4	40	—	0.1 ^q	20 ^b	—	—	—	4 ^q
HS5010N	HS	B	2	—	1.22	2	3	—	0.1 ⁿ	5 ^b	5 ^f	—	—	0.05 ^a
ICL8069A	IL	B	2	—	1.23	2	10	—	0.5	10 ^b	—	—	—	0.2 ^a
TSC9491	TS	B	2	—	1.22	2	30	—	0.1 ^k	0.5 ^b	—	—	—	1.2 ^k
Three-terminal type														
REF-01A	PM	B	8	•	10.0	0.3	3	12	1	10	20	—	0.006	0.005
REF-02A	PM	B	8	•	5.0	0.3	3	7	1	10	10	—	0.006	0.005
REF-03E	PM	B	8	•	2.5	0.3	3	4.5	1	10	5	—	0.006	0.05
REF-05	PM	B	8	•	5.0	0.3	3	7	1	10	10	100 ^m	0.006	0.05
REF-08G	PM	Z	8	•	-10.0	0.2	10 ^m	-11.4	2 ^m	10	10	—	0.02 ^m	0.2 ^m
REF-10	PM	B	8	•	10.0	0.3	3	12	1	10	20	50 ^m	0.006	0.05
REF10KM	BB	Z	8	•	10.0	0.05	1 ^m	13.5	4.5	10	6	10	0.001	0.01
REF-43E	PM	B	8	•	2.5	0.05	3 ^m	4.5	0.2 ^m	10	8 ^{gm}	—	0.0002 ^m	0.03 ^m
LH0070-1	NS	Z	3	—	10.0	0.1	4	12.5	3	10	20	—	0.001	0.01
REF101KM	BB	Z	8	•	10.0	0.05	1 ^m	13.5	4.5	10	6	25	0.0003	0.003
LM368Y-2.5	NS	B	8	•	2.5	0.2	11	4.9	0.35	10	12	—	0.0001	0.003
LM368-5	NS	B	4	•	5.0	0.1	15	7.5	0.25	10	16	—	0.0001	0.003
LM368-10	NS	B	4	•	10.0	0.1	15	12.5	0.25	10	30	—	0.0001	0.003
LM369B	NS	Z	3,8	•	10.0	0.05	1.5	13	1.4	10	4	6	0.0002	0.003
AD580M	AD	B	3	—	2.5	1	10	4.5	1	10	60	25	0.04	0.4

Type	Mfg ^a	B ^b	gap/zener	Terminals	Trim	Voltage (V)	Acc'y (%)	Tempco typ (ppm/°C)	Min supply voltage (V)	Supply curr (mA)	Output curr max (mA)	Noise voltage 0.1-10Hz typ (μV pp)	Long-term stability typ (ppm/1000h)	Regulation	
														Line typ (%/V)	Load 0-10mA typ (%)
Three-terminal type (cont'd)															
AD581L	AD+	B	3	—	10.0	0.05	5	12	0.75	10	50	25	0.005	0.002	
AD584L	AD	B	8	•	2.5	0.05	10	5	0.75	18	50	25	0.005	0.002	
"	AD				5.0	0.06	5	7.5	0.75	15	50	25	0.005	0.002	
"	AD				7.5	0.06	5	10	0.75	13	50	25	0.005	0.002	
"	AD				10.0	0.1	5	12.5	0.75	10	50	25	0.005	0.002	
AD586L	AD	Z	8	•	5.0	0.05	5 ^m	—	5 ^m	10	—	15	—	—	
AD587L	AD	Z	8	•	10.0	0.05	5 ^m	—	5 ^m	10	—	15	—	—	
AD588B	AD	Z	14	•	±10.0	0.01	1.5 ^m	±14	±10	±10	10	25 ^m	0.002 ^m	0.01 ^m	
MAX671C	MA	Z	14	•	10.0	0.01	1 ^m	13.5	9	10	12	50	0.005 ^m	0.01 ^m	
AD689L	AD	Z	8	•	8.192	0.05	5 ^m	10.8	2	±10	2	15	0.002 ^m	0.01 ^m	
R675C-3	HS	Z	14	•	±10.0	0.05	5	±13	+15,-3 ^m	10	—	—	0.003 ^m	0.02 ^m	
LT1019A-2.5	LT	B	8	•	2.5	0.002	3	4	0.7	10	6	—	0.00005	0.008	
LT1021B-5	LT	Z	8	•	5.0	1	2	7	0.8	10	3	15	0.0004	0.01	
LT1031B	LT	Z	3	—	10.0	0.05	3	11	1.2	10	6	15	0.00005	0.01	
MC1403A	MO	B	8	—	2.5	1	10	4.5	1.2	10	—	—	0.002	0.06	
MC1404AU5	MO	B	8	•	5.0	1	10	7.5	1.2	10	12	25	0.001	0.06	
MC1404AU10	MO	B	8	•	10.0	1	10	12.5	1.2	10	12	25	0.0006	0.06	
AD2702L ⁱ	AD+	Z	14	•	±10.0	0.05	5 ^m	±13	+12,-2	±10	50	100	0.03 ^m	0.05 ^m	
AD2712L ⁱ	AD+	Z	14	•	±10.0	0.01	1 ^m	±13	+12,-2	±5	30	25	0.013	0.003 ^j	
LP2950ACZ	NS	B	3	—	5.0	0.5	20	5.4	0.08	100	—	—	0.002	0.004	
ICL8212	IL	B	8	•	1.15	3	200	1.8	0.035	20	—	—	0.2	—	
TSC9495	TS	B	8	•	5.0	1	20	7	1	8	12	—	0.01	0.06	
TSC9496	TS	B	8	•	10.0	1	20	12	1	8	25	—	0.01	0.06	

(a) 0 to 1mA. (b) max zener curr. (c) on-chip heater/thermostat. (d) specified for 10μA to 20mA operating curr.
(e) 1Hz to 10Hz. (f) 10Hz to 10kHz, rms. (g) 10Hz to 1kHz, rms. (h) spec'd for 50μA to 5mA. (i) 2700,2710: +10V;
2701: -10V; 2702,2712: ±10V. (j) 0 to 5mA. (k) spec'd for 50μA to 500μA. (l) spec'd for 0.5 to 20mA. (m) min or max.
(n) 1 to 20mA, max. (o) specified for 0.5 to 10mA. (p) specified for 20μA to 20mA. (q) specified for 100μA to 20mA.
(r) specified for 1mA to 5mA.

total voltage equals the silicon bandgap voltage (extrapolated to absolute zero), about 1.22 volts. The circuit in the box is the reference. Its own output is used (via R_3) to create the constant current we initially assumed.

Figure 6.25 shows another very popular bandgap reference circuit (it replaces the components in the box in Figure 6.24). Q_1 and Q_2 are a matched pair, forced to operate at a ratio of emitter currents of 10:1 by feedback from the collector voltages. The difference in V_{BE} s is $(kT/q) \log_e 10$,

making Q_2 's emitter current proportional to T (the preceding voltage applied across R_1). But since Q_1 's collector current is larger by a factor of 10, it also is proportional to T . Thus, the total emitter current is proportional to T , and therefore it generates a positive-tempco voltage across R_2 . That voltage can be used as a thermometer output, by the way, as will be discussed shortly. R_2 's voltage is added to Q_1 's V_{BE} to generate a stable reference of zero tempco at the base. Bandgap references appear in many variations, but they

all feature the summation of V_{BE} with a voltage generated from a pair of transistors operated with some ratio of current densities.

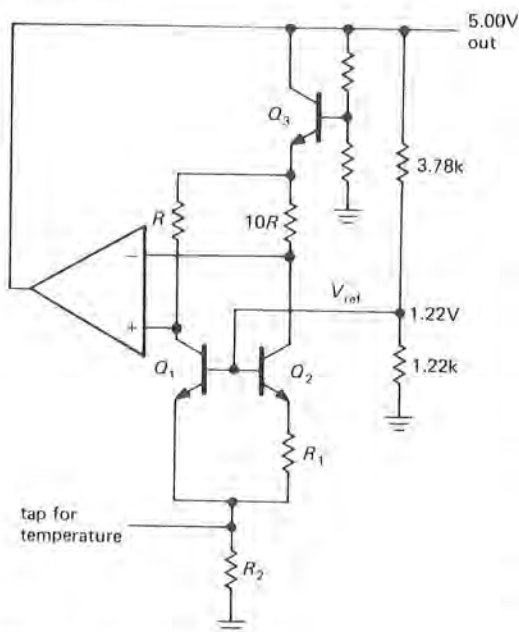


Figure 6.25

□ IC bandgap references

An example of an IC bandgap reference is the inexpensive 2-terminal LM385-1.2, with a nominal operating voltage of 1.235 volts, $\pm 1\%$ (the companion LM385-2.5 uses internal circuitry to generate 2.50V), usable down to $10\mu\text{A}$. That's much less than you can run any zener at, making these references excellent for micropower equipment (see Chapter 14). The low reference voltage (1.235V) is often much more convenient than the approximately 5 volt minimum usable voltage for zeners (you can get zeners rated at voltages as low as 3.3V, but they are pretty awful, with very soft knees). The best grade of LM385 guarantees $30\text{ppm}/^\circ\text{C}$ maximum tempco and has a typical dynamic impedance of 1 ohm at $100\mu\text{A}$. Compare this with the equivalent figures for a 1N4370 2.4 volt

zener diode: tempco $800\text{ppm}/^\circ\text{C}$ (typ), dynamic impedance ≈ 3000 ohms at $100\mu\text{A}$, at which the "zener voltage" (specified as 2.4V at 20mA) is about 1.1 volt! When you need a precision stable voltage reference, these excellent bandgap ICs put conventional zener diodes to shame.

If you're willing to spend a bit more money, you can find bandgap references of excellent stability, for example, the 2-terminal LT1029, or the 3-terminal REF-43 (2.50V, $3\text{ppm}/^\circ\text{C}$ max). The latter type, like the 3-terminal references based on zener technology, requires a dc supply. Table 6.7 lists most available bandgap (and zener) references, both 2-terminal and 3-terminal.

One other interesting voltage reference is the TL431C. It is an inexpensive "programmable zener" reference, and it is used as shown in Figure 6.26. The "zener" (made from a V_{BE} circuit) turns on when the control voltage reaches 2.75 volts; the device draws only a few microamps from the control terminal and gives a typical tempco of output voltage of $10\text{ppm}/^\circ\text{C}$. The circuit values shown give a zener voltage of 10.0 volts, for example. This device comes in a mini-DIP package and can handle currents to 100mA.

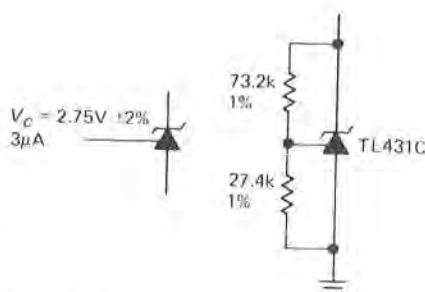


Figure 6.26

□ Bandgap temperature sensors

The predictable V_{BE} variation with temperature can be exploited to make a temperature-measuring IC. The REF-02, for

instance, generates an additional output voltage that varies linearly with temperature (see preceding discussion). With some simple external circuitry you can generate an output voltage that tells you the chip temperature, accurate to 1% over the full "military" temperature range (-55°C to $+125^{\circ}\text{C}$). The AD590, intended for temperature measurement only, generates an accurate current of $1\mu\text{A}/^{\circ}\text{K}$. It's a 2-terminal device; you just put a voltage across it (4–30V) and measure the current. The LM334 can also be used in this manner. Other sensors, such as the LM35 and LM335, generate accurate voltage outputs with a slope of $+10\text{mV}/^{\circ}\text{C}$. Section 15.01 has a detailed discussion on all these temperature "transducers."

Three-terminal precision references

As we remarked earlier, it is possible to make voltage references of remarkable temperature stability (down to $1\text{ppm}/^{\circ}\text{C}$ or less). This is particularly impressive when you consider that the venerable Weston cell, the traditional voltage reference through the ages, has a temperature coefficient of $40\text{ppm}/^{\circ}\text{C}$ (see Section 15.11). There are two techniques used to make such references.

- 1. *Temperature-stabilized references.* A good approach to achieving excellent temperature stability in a voltage reference circuit (or any other circuit, for that matter) is to hold the reference, and perhaps its associated electronics, at a constant elevated temperature. You will see simple techniques for doing this in Chapter 15 (one obvious method is to use a bandgap temperature sensor to control a heater). In this way the circuit can deliver equivalent performance with a greatly relaxed temperature coefficient, since the actual circuit components are isolated from external temperature fluctuations. Of greater interest for precision circuitry is the

ability to deliver significantly improved performance by putting an already well-compensated reference circuit into a constant-temperature environment.

This technique of temperature-stabilized or "ovenized" circuits has been used for many years, particularly for ultrastable oscillator circuits. There are commercially available power supplies and precision voltage references that use ovenized reference circuits. This method works well, but it has the drawbacks of bulkiness, relatively large heater power consumption, and sluggish warm-up (typically 10min or more). These problems are effectively eliminated if the thermal stabilization is done at the chip level by integrating a heater circuit (with sensor) onto the integrated circuit itself. This approach was pioneered in the 1960s by Fairchild with the $\mu\text{A}726$ and $\mu\text{A}727$ temperature-stabilized differential pair and preamp, respectively.

More recently, temperature-stabilized voltage references such as the National LM199 series have appeared. It offers a temperature coefficient of $0.00002\%/^{\circ}\text{C}$ (typ), which is a mere $0.2\text{ppm}/^{\circ}\text{C}$. These references are packaged in standard metal transistor cans (TO-46); they consume about 0.25 watt of heater power and come up to temperature in 3 seconds. Users should be aware that the subsequent op-amp circuitry, and even precision wirewound resistors with their $\pm 2.5\text{ppm}/^{\circ}\text{C}$ tempco, may degrade performance considerably, unless extreme care is used in design. In particular, low-drift precision op-amps such as the OP-07, with $0.2\mu\text{V}/^{\circ}\text{C}$ (typ) input-stage drift, are essential. These aspects of precision circuit design are discussed in Sections 7.01 to 7.06.

One caution when using the LM399: The chip can be damaged if the heater supply hovers below 7.5 volts for any length of time.

The LT1019 bandgap reference, though normally operated unheated, has an on-chip heater and temperature sensor. So

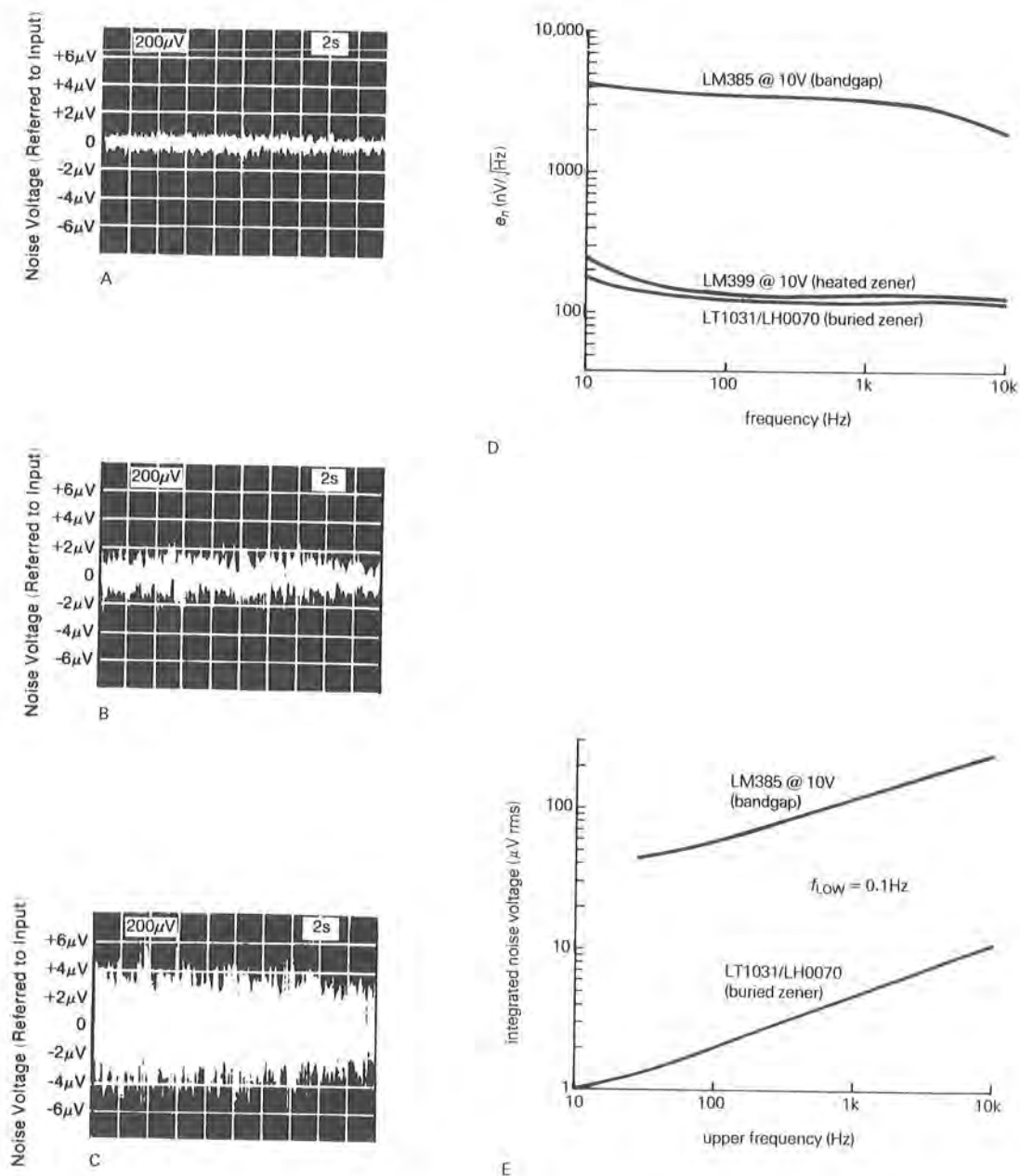


Figure 6.27. Buried zener references (A) have lower noise than either heated zeners (B) or bandgap references (C). (Courtesy of Burr-Brown Corporation.) (D) Noise density (e_n) comparison; (E) integrated noise voltage comparison.

you can use it like the LM399, to get tempcos less than 2ppm/°C. However, unlike the LM399, the LT1019 requires some external circuitry to implement the thermostat (an op-amp and a half dozen components).

- 2. *Precision unheated references.* The thermostated LM399 has excellent tempco, but it does not exhibit extraordinary noise or long-term drift specs (see Table 6.7). The chip also takes a few seconds to heat up, and it uses plenty of power (4W at start-up, 250mW stabilized).

Clever chip design has made possible unheated references of equivalent stability. The REF10KM and REF101KM from Burr-Brown have tempcos of 1ppm/°C (max), with no heater power or warm-up delays. Furthermore, they exhibit lower long-term drift and noise than the LM399-style references. Other 3-terminal references with 1ppm/°C maximum tempco are the MAX671 from Maxim and the AD2710/2712 references from Analog Devices. In 2-terminal configurations the only contender is the magnificent LTZ1000 from Linear Technology, with its claimed 0.05ppm/°C tempco. It also claims long-term drift and noise specs that are a factor of 10 better than any other reference of any kind. The LTZ1000 does require a good external biasing circuit, which you can make with an op-amp and a few parts. All of these high stability references (including the heated LM399) use buried zeners, which additionally provide much lower noise than ordinary zener or band-gap references (Figure 6.27).

THREE-TERMINAL AND FOUR-TERMINAL REGULATORS

6.16 Three-terminal regulators

For most noncritical applications the best choice for a voltage regulator is the simple 3-terminal type. It has only three connections (input, output, and ground) and

is factory-trimmed to provide a fixed output. Typical of this type is the 78xx. The voltage is specified by the last two digits of the part number and can be any of the following: 05, 06, 08, 10, 12, 15, 18, or 24. Figure 6.28 shows how easy it is to make a +5 volt regulator, for instance, with one of these regulators. The capacitor across the output improves transient response and keeps the impedance low at high frequencies (an input capacitor of at least 0.33μF should be used in addition if the regulator is located a considerable distance from the filter capacitors). The 7800 series is available in plastic or metal power packages (same as power transistors). A low-power version, the 78Lxx, comes in the same plastic and metal packages as small-signal transistors (see Table 6.8). The 7900 series of negative regulators works the same way (with negative input voltage, of course). The 7800 series can provide up to 1 amp load current and has on-chip circuitry to prevent damage in the event of overheating or excessive load current; the chip simply shuts down, rather than blowing out. In addition, on-chip circuitry prevents operation outside the transistor safe operating area (see Section 6.07) by reducing available output current for large input-output voltage differential. These regulators are inexpensive and easy to use, and they make it practical to design a system with many printed-circuit boards in which the unregulated dc is brought to each board and regulation is done locally on each circuit card.

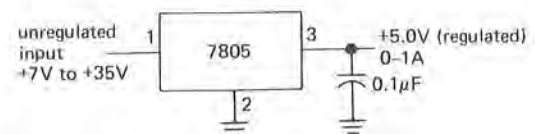


Figure 6.28

Three-terminal fixed regulators come in some highly useful variants. The LP2950 works just like a 7805, but draws only

TABLE 6.8. FIXED VOLTAGE REGULATORS

Type	Pkg	V _{out} (V)	Accuracy (%)	Output current (max) ^a						Input voltage	
				@75°C case	No heatsink ^b			Regulation (typ)			
					I _{out} (A)	I _{out} (A)	P _{diss} (W)	Load ^c (mV)	Line ^d (mV)	Θ _{JC} (°C/W)	min ⁱ (V)
Positive											
LM2950CZ-5.0	TO-92	5	1	0.08	0.1	0.5	2	1.5	160	5.4	30
LM2931Z-5.0	TO-92	5	5	0.1	0.1	0.5	14	3	160	5.3	26
LM78L05ACZ	TO-92	5	4	0.1	0.1	0.6	5	50	160	7	35
LM330T-5.0 ^g	TO-220	5	4	0.15	0.15	1.5	14	20	4	5.3	26
TL750L05	TO-92	5	4	0.15	0.15	0.6	20	6	160	5.6	26
LM2984CT	TO-220 ^h	5	3	0.5	0.5	2	12	4	3	5.5	26
LM2925T	TO-220	5	5	0.75	0.5	2	10	8	3	5.6	26
LM2935T	TO-220	5	5	0.75	0.5	2	10	8	3	5.5	26
LM309K	TO-3	5	4	1	0.6	2.2	20	4	3	7	35
LT1005CT	TO-220	5	2	1	0.5	2	5	5	3	7	20
LM2940T-5.0	TO-220	5	3	1	0.5	2	35	20	3	5.5	26
LM7805CK	TO-3	5	4	1	0.6	2.2	10	3	3.5	7	35
LM7805CT	TO-220	5	4	1	0.45	1.7	10	3	3	7	35
LM7815CT	TO-220	15	4	1	0.15	1.7	12	4	3	17	35
LT1086-5CT	TO-220	5	1	1.5	0.5	2	5	0.5	3	6.3	30
LAS16A05	TO-3	5	2	2	0.75	2.8	30 ^m	100 ^m	2.5	7.6	30
LM323K	TO-3	5	4	3	0.6	2	25	5	2	7	20
LT1035CK	TO-3	5	2	3	0.8	3	10	5	1.5	7.3	20
LT1085-5CT	TO-220	5	1	3	0.5	2	5	0.5	3	6.3	30
LAS14A05	TO-3	5	2	3	0.8	3	30 ^m	50 ^m	2.3	7.5	35
LT1003CK	TO-3	5	2	5	0.8	3	25	5	1	7.3	20
LT1084-5CK	TO-3	5	1	5	0.8	3	5	0.5	1.6	6.3	30
LAS19A05	TO-3	5	2	5	0.8	3	30 ^m	50 ^m	0.9	7.6	30
LT1083-5CK	TO-3	5	1	7.5	0.8	3	5	0.5	1.6	6.3	30
LAS3905	TO-3	5	5	8	0.8	3	20 ^m	100 ^m	0.7	7.6	25
Negative											
LM79L15ACZ	TO-92	-15	4	0.1	0.05	0.6	75 ^m	45 ^m	160	-17	-35
LM7915CK	TO-3	-15	4	1	0.2	2.2	4	3	3.5	-16.5	-35
LM7915CT	TO-220	-15	4	1	0.15	1.7	4	3	3	-16.5	-35
LM345K-5.0	TO-3	-5	4	3	0.2	2.1	10	5	2	-7.5	-20

(a) with $V_{in}=1.75V_{out}$. (b) 50°C ambient. (c) 0 to I_{max} . (d) $\Delta V_{in}=15V$. (e) ΔV_{out} for 0°C to 100°C junc temp.
 (f) 1000 hours. (g) similar to LM2930T-5.0, LM2931T-5.0. (h) wide TO-220. (i) at I_{max} . (m) min or max.
 (l) typical. All include internal thermal shutdown and current-limiting circuitry. Most are available in ± 5 , 6, 8, 10, 12, 15, 18, and 24V units; a few are available in -2, -3, -4, -5.2, -9, +2.6, +9, and +17V units.

Type	120Hz ripple reject typ (dB)	Temp stab ^e typ (mV)	Long- term stab ^f max (%)	Output impedance		Comments
				10Hz (Ω)	10kHz (Ω)	
LM2950CZ-5.0	70	10	—	0.01	0.5	micropower, 1%
LM2931Z-5.0	80	—	0.4 ^l	0.1	0.2	low dropout, low power
LM78L05ACZ	50	—	0.25	0.2	0.2	small; LM240LAZ-5.0
LM330T-5.0 ^g	56	25	0.4 ^l	0.1	0.2	low dropout; 2930
TL750L05	65	50	—	—	—	TL751 has enable
LM2984CT	70	3	0.4 ^l	0.01	0.02	dual outputs (μ P); reset, on/off
LM2925T	66	—	0.4 ^l	0.2	0.2	microprocessor; reset
LM2935T	66	—	0.4 ^l	0.02	0.02	dual outputs (μ P); reset, on/off
LM309K	80	50	0.4	0.04	0.05	original +5V regulator
LT1005CT	70	25	—	0.003	0.01	dual outputs (μ P)
LM2940T-5.0	72	20	0.4 ^l	0.03	0.03	
LM7805CK	80	30	0.4	0.01	0.03	LM340K-5
LM7805CT	80	30	0.4	0.01	0.03	popular; LM340T-5
LM7815CT	70	100	0.4	0.02	0.05	LM340T-15
LT1086-5CT	63	25	1	—	—	low dropout
LAS16A05	75	—	—	0.002	0.02	Lambda, monolithic
LM323K	70	30	0.7	0.01	0.02	
LT1035CK	70	25	—	0.003	0.01	dual +5; 1036 is +12/+5
LT1085-5CT	63	25	1	—	—	low dropout
LAS14A05	70	100 ^m	—	0.001	0.003	Lambda, monolithic
LT1003CK	66	25	0.7	0.003	0.02	
LT1084-5CK	63	25	1	—	—	low dropout
LAS19A05	70	150 ^m	—	0.01	0.2	lambda, monolithic
LT1083-5CK	63	25	1	—	—	low dropout
LAS3905	60 ^m	100	—	0.004	0.01	Lambda, monolithic
LM79L15ACZ	40	—	0.4 ^l	0.05	0.05	small; LM320LZ-15
LM7915CK	60	60	0.4	0.06	0.07	LM320KC-15
LM7915CT	60	60	0.4	0.06	0.07	LM320T-15
LM345K-5.0	65	25	1.0	0.02	0.04	

75 μ A of quiescent current (compared with the 7805's 5mA, or the 78L05's 3mA); it also regulates with as little as a 0.4 volt drop from unregulated input to regulated output (called the "dropout voltage"), compared with 2 volts dropout for the classic

7805. The LM2931 is also low-dropout, but you might call it *millipower* (0.4mA quiescent current), compared with the "micropower" LP2950. Low-dropout regulators also come in high-current versions — for example, the LT1085/4/3 series from

LTC (3A, 5A, and 7.5A, respectively, with both +5V and +12V available in each type). Regulators like the LM2984 are basically 3-terminal fixed regulators, but with extra outputs to signal a microprocessor that power has failed, or resumed. Finally, regulators like the 4195 contain a pair of 3-terminal 15 volt regulators, one positive and one negative. We'll say a bit more about these special regulators shortly.

6.17 Three-terminal adjustable regulators

Sometimes you want a nonstandard regulated voltage (say +9V, to emulate a battery) and can't use a 78xx-type fixed regulator. Or perhaps you want a standard voltage, but set more accurately than the $\pm 3\%$ accuracy typical of fixed regulators. By now you're spoiled by the simplicity of 3-terminal fixed regulators, and therefore you can't imagine using a 723-type regulator circuit, with all its required external components. What to do? Get an "adjustable 3-terminal regulator"! Table 6.8 lists the characteristics of a representative selection of 3-terminal fixed regulators.

These wonderful ICs are typified by the classic LM317 from National. This regulator has no ground terminal; instead, it adjusts V_{out} to maintain a constant 1.25 volts (bandgap) from the output terminal to the "adjustment" terminal. Figure 6.29 shows the easiest way to use it. The regulator puts 1.25 volts across R_1 , so 5mA flows through it. The adjustment terminal draws very little current (50–100 μ A), so the output voltage is just

$$V_{out} = 1.25(1 + R_2/R_1) \text{ volts}$$

In this case the output voltage is adjustable from 1.25 volts to 25 volts. For a fixed-output-voltage application, R_2 will normally be adjustable only over a narrow range, to improve settability (use a fixed resistor in series with a trimmer). Choose your resistive divider values low enough

to allow for a 50 μ A change in adjustment current with temperature. Because the loop compensation for the regulator is the output capacitor, larger values must be used compared with other designs. At least a 1 μ F tantalum is required, but we recommend something more like 6.8 μ F.

The 317 is available in several packages, including the plastic power package (TO-220), the metal power package (TO-3), and the small transistor packages (metal, TO-5; plastic, TO-92). In the power packages it can deliver up to 1.5 amps, with proper heat sinking. Because it doesn't "see" ground, it can be used for high-voltage regulators, as long as the input-output voltage differential doesn't exceed the rated maximum of 40 volts (60V for the LM317HV high-voltage variant).

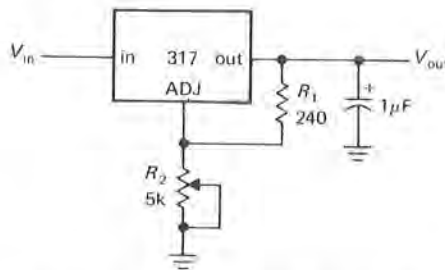


Figure 6.29. Three-terminal adjustable regulator.

EXERCISE 6.5

Design a +5 volt regulator with the 317. Provide $\pm 20\%$ voltage adjustment range with a trimmer pot.

Three-terminal adjustable regulators are available with higher current ratings, e.g., the LM350 (3A), the LM338 (5A), and the LM396 (10A), and also with higher voltage ratings, e.g., the LM317H (60V) and the TL783 (125V). Read the data sheets carefully before using these parts, noting bypass capacitor requirements and safety diode suggestions. As with the fixed

3-terminal regulators, you can get low-dropout versions (e.g., the LT1085, with 1.3V dropout at 3.5A), and you can get micropower versions (e.g., the LP2951, the adjustable variant of the fixed 5V LP2950; both have $I_Q = 75\mu\text{A}$). You can also get *negative* versions, although there's less variety: The LM337 is the negative cousin of the LM317 (1.5A), and the LM333 is a negative LM350 (3A).

□ Four-terminal regulators

Three-terminal adjustable regulators are the favorite for noncritical requirements. Historically they were preceded by four-terminal adjustable regulators, which you connect as shown in Figure 6.30. You drive the "control" terminal with a sample of the output; the regulator adjusts the output to keep the control terminal at a fixed voltage (+3.8V for the Lambda regulators in Table 6.9, +5V for the μA79G , and -2.2V for the negative regulators). Four-terminal regulators aren't any better than the simpler 3-terminal variety (but they aren't any worse, either), and we mention them here for completeness.

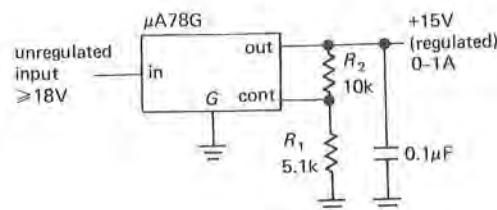


Figure 6.30

6.18 Additional comments about 3-terminal regulators

General characteristics of 3- and 4-terminal regulators

The following specifications are typical for most 3- and 4-terminal regulators, both fixed and adjustable, and they may be

useful as a rough guide to the performance you can expect:

Output voltage tolerance:	1–2%
Dropout voltage:	0.5–2 volts
Maximum input voltage:	35 volts (except TL783 to +125V)
Ripple rejection:	0.01–0.1%
Spike rejection:	0.1–0.3%
Load regulation:	0.1–0.5%, full load change
dc input rejection:	0.2%
Temperature stability:	0.5%, over full temp range

Improving ripple rejection

The circuit of Figure 6.29 is the standard 3-terminal regulator, and it works fine. However, the addition of a 10 μF bypass capacitor from the adjust (ADJ) terminal to ground (Fig. 6.31) improves the ripple (and spike) rejection by about 15dB (factor of 5 in voltage). For example, the LM317 ripple rejection factor goes from 65dB to 80dB (the latter is 0.1mV output ripple when supplied with 1V input ripple, a typical value). Be sure to include the safety discharge diode; look at the specification sheet of the particular regulator for more details.

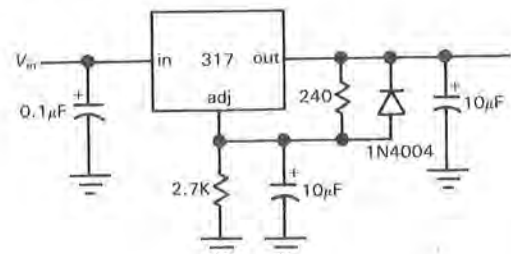


Figure 6.31. The ADJ pin may be bypassed for lower noise and ripple, but a safety discharge diode must be included.

Low-dropout regulators

As we mentioned earlier, most series regulators need at least 2 volts of "headroom" to function; that's because the base of the *n*pn pass transistor is a V_{BE} drop above

TABLE 6.9. ADJUSTABLE VOLTAGE REGULATORS

Type	Polarity	Output voltage		Regulation (typ)			Input voltage		Dropout voltage @I _{max}	120Hz ripple reject typ (dB)	Temp stab ^c		Long-term stab ^d		Output impedance		Therm lim	Curr lim	Comments
		min (V)	max (V)	I _{max} (A)	Load ^a (%)	Line ^b (%)	θ _{JC} (°C/W)	min (V)			max (V)	typ (%)	max (%)	10Hz (Ω)	10kHz (Ω)				
<i>Three-terminal</i>																			
LM317L	+	TO-92	1.2	37	0.1	0.1	0.15	160 ^h	2.5 ^t	65	0.5	1	0.07	4	•	miniature			
LM337L	-	TO-92	1.2	37	0.1	0.1	0.15	160 ^h	2.5 ^t	65	0.5	-	-	-	•	miniature (neg 317L)			
LM317H	+	TO-39	1.2	37	0.5	0.1	0.2	12	2 ^t	80	0.6	0.3	0.01	0.03	•	317 in TO-39			
LM337H	-	TO-39	1.2	37	0.5	0.3	0.2	12	2 ^t	75	0.5	0.3	0.02	0.02	•	negative 317H			
TL783C	+	TO-220	1.3	125	0.7	0.2 ^f	0.02	4	10	50	0.3	0.2	0.05	0.3	•	MOSFET, high voltage			
LM317T	+	TO-220	1.2	37	1.5	0.1	0.2	4	2.5 ^t	80	0.6	0.3	0.01	0.03	•	popular			
LM317HVK	+	TO-3	1.2	57	1.5	0.1	0.2	2.3	2.5 ^t	80	0.6	0.3	0.01	0.03	•	high voltage 317			
LM337T	-	TO-220	1.2	37	1.5	0.3	0.2	4	2.5 ^t	75	0.5	0.3	0.02	0.02	•	negative 317			
LM337HVK	-	TO-3	1.2	47	1.5	0.3	0.2	2.3	2.5 ^t	75	0.5	0.3	0.02	0.02	•	high voltage 337			
LT1086CP	+	TO-220	1.3	30	1.5	0.1	0.02	-	1.5	75	0.5	1	-	-	•	low dropout			
LM350K	+	TO-3	1.2	32	3	0.1	0.1	2	2.5 ^t	80	0.6	0.3	0.005	0.02	•	3A monolithic			
IP3R07T	+	TO-220	1.2	37	3	0.1	0.08	2.3	0.8 ^t	65	-	-	-	-	•	two unreg inputs			
LM333T	-	TO-220	1.2	32	3	0.2	0.02	50	2.5 ^t	60	0.5	0.2	-	-	•	neg 350; LT1033 is imprvd			
LT1085CT	+	TO-220	1.3	30	3	0.1	0.02	3	1.5	75	0.5	1	-	-	•	low dropout			
LM338K	+	TO-3	1.2	32	5	0.1	0.1	2	2.5 ^t	80	0.6	0.3	-	-	•	5A monolithic			
LT1084CP	+	TO-247	1.3	30	5	0.1	0.02	2.3	1.5	75	0.5	1	-	-	•	low dropout			
LT1083CP	+	TO-247	1.3	30	7.5	0.1	0.02	1.6	1.5	75	0.5	1	-	-	•	low dropout			
LM396K	+	TO-3	1.2	15	10	0.4 ^m	0.08	1	2.1 ^t	74	0.3	1	0.01	0.02	•	10A monolithic			
LT1038CK	+	TO-3	1.2	32	10	0.1	0.08	1	2.5 ^t	60	1	1	0.005	0.1	•	10A monolithic, 1% acc'y			

Four-terminal															
μ A78GU1C	+	TO-220	5	30	1	1 ^m	1 ^m	7.5	7.5	40	2.5	80	3 ^m	•	TO-39 pkg avail
μ A79GU1C	-	TO-220	-2.5	-30	1	1 ^m	1 ^m	7.5	-7	-40	2 ^t	60	3 ^m	•	TO-39 pkg avail
LAS15U	+	TO-3	4	30	1.5	0.6 ^m	2 ^m	3	6.5	40	2.4	70	3 ^m	•	Lambda
LAS18U	-	TO-3	-2.6	-30	1.5	0.6 ^m	2 ^m	3	-5	-40	2.1	60	3 ^m	•	Lambda
LAS16U	+	TO-3	4	30	2	0.6 ^m	2 ^m	2.5	6.5	35	2.6	70	2 ^m	•	Lambda
LAS14AU	+	TO-3	4	35	3	0.6 ^m	1 ^m	1.5	6.5	40	2.3	70	2 ^m	•	Lambda
LAS19U	+	TO-3	4	30	5	0.6 ^m	1 ^m	0.9	6.5	35	2.6	65	2 ^m	•	Lambda
LAS39U	+	TO-3	4	16	8	0.6 ^m	2 ^m	0.7	6.6	25	2.6	60 ^m	3 ^m	•	Lambda
Multiterminal															
LM376N	+	DIP-8	5	37	0.03	0.2 ^m	0.6 ^m	190 ^h	9	40	3	60 ^m	1 ^m	-	E
LM304H	-	TO-5	0	-40	0.03	1mV	0.2	45	-8	-40	2	65	0.3	-	E
ICL7663S	+	DIP-8	1.3	16	0.04	0.4 ^t	0.5	200 ^h	1.5	16	1 ^t	20	1	-	E
MAX664	-	DIP-8	-1.3	-16	0.04	0.8 ^t	0.5	120 ^h	-2	-16	0.3 ^t	15	1	-	E
LM305AH	+	TO-5	4.5	40	0.05	0.03	0.3	45	8.5	50	3	80	0.3	-	E
LM2931CT	+	TO-220	3	24	0.1	0.3	0.06 ^f	3	3.6	26	0.3 ^t	60 ^f	-	•	low dropout, low power
LP2951CN	+	DIP-8	1.3	29	0.1	0.1 ^t	0.03 ^f	105	1.7	30	0.4 ^t	70 ^f	0.5	•	low dropout, micropower
LT1020CN	+	DIP-14	2.5	35	0.13	0.2	0.15	60 ^h	4.5	36	0.4 ^t	60	1	•	micropower
NE550N	+	DIP-14	2	40	0.15	0.03	0.08	150 ^h	8.5	40	3	90	0.2	0.1 ^t	E
μ A723PC	+	DIP-14	2	37	0.15	0.03	0.1	150 ^h	9.5	40	3	75	0.3	0.1	E
LAS1000	+	TO-5	3	38	0.15	0.1 ^m	0.2	150 ^h	5	40	2	60 ^m	1.5 ^m	•	Lambda, improved 723
LAS1100	+	TO-5	3	48	0.15	0.1 ^m	0.2	150 ^h	5	50	2	60 ^m	1.5 ^m	•	E high voltage LAS1000
SG3532J	+	DIP-14	2	38	0.17	0.1	0.1	125 ^h	4.7	40	2	66	0.5	•	E improved 723
MC1469R	+	TO-66	2.5	32	0.6	0.005	0.05	7	9	35	3	100	0.2	-	E precision, may oscillate
MC1463R	-	TO-66	-3.8	-32	0.6	0.005	0.05	17	-9	-35	3	90	0.2	-	E neg MC1469
LM2941CT	+	TO-220	1.3	25	1			3	26	1	74	0.4	0.4 ^t	-	low dropout
LAS2200	+	module	2.5	28	5	0.2 ^m	0.15 ^m	2	9.6	40	2.5	60 ^m	0.7 ^m	•	•
LAS3000	+	module	2.7	29	10	0.2 ^m	0.15 ^m	1.3	7.9	40	2.5	60 ^m	1.5 ^m	•	Lambda hybrid; 2 unreg inputs
LAS5000	+	module	4.8	29	20	0.2 ^m	0.2 ^m	0.7	11.9	40	2.5	60 ^m	1.5 ^m	•	Lambda hybrid; 2 unreg inputs
LAS7000	+	module	4.8	29	30	0.2 ^m	0.2 ^m	0.4	12.3	40	2.5	60 ^m	1.5 ^m	•	Lambda hybrid; 2 unreg inputs
MC1466L	+	DIP-14	0	1000	-	0.02	0.05	170 ^h	-	-	2 ^t	70	0.4	-	E lab supply; good curr lim
LAS3700	+	TO-5	0	1000	-	0.003	0.15	220	-	-	-	65	0.5 ^m	-	E floating reg with on-chip heater
(a) 10% to 50% I _{max} . (b) for $\Delta V_{in}=15V$. (c) ΔV_{out} for 0°C to 100°C T _J . (d) 1000 hrs. (e) max V _{in} -V _{out} . (f) at 5V. (h) Θ_{JA} . (i) E - external; I - internal. (m) min/max. (t) typ.															

the output, and it has to be driven by a driver transistor, usually another *nnp* whose base is pulled up with a current mirror. That's already two V_{BE} drops. Furthermore, you need to allow another V_{BE} drop across the current-sensing resistor for short-circuit protection; see Figure 6.32A, a simplified schematic of the 78Lxx. The three V_{BE} s add up to about 2 volts, below

which the regulator drops out of regulation at full current.

By using a *pnp* (or *p*-channel MOSFET) pass transistor, the dropout voltage can be reduced from the three V_{BE} drop of the conventional *nnp* scheme, down nearly to the transistor saturation voltage. Figure 6.32B shows a simplified schematic of the LM330 low-dropout fixed +5 volt

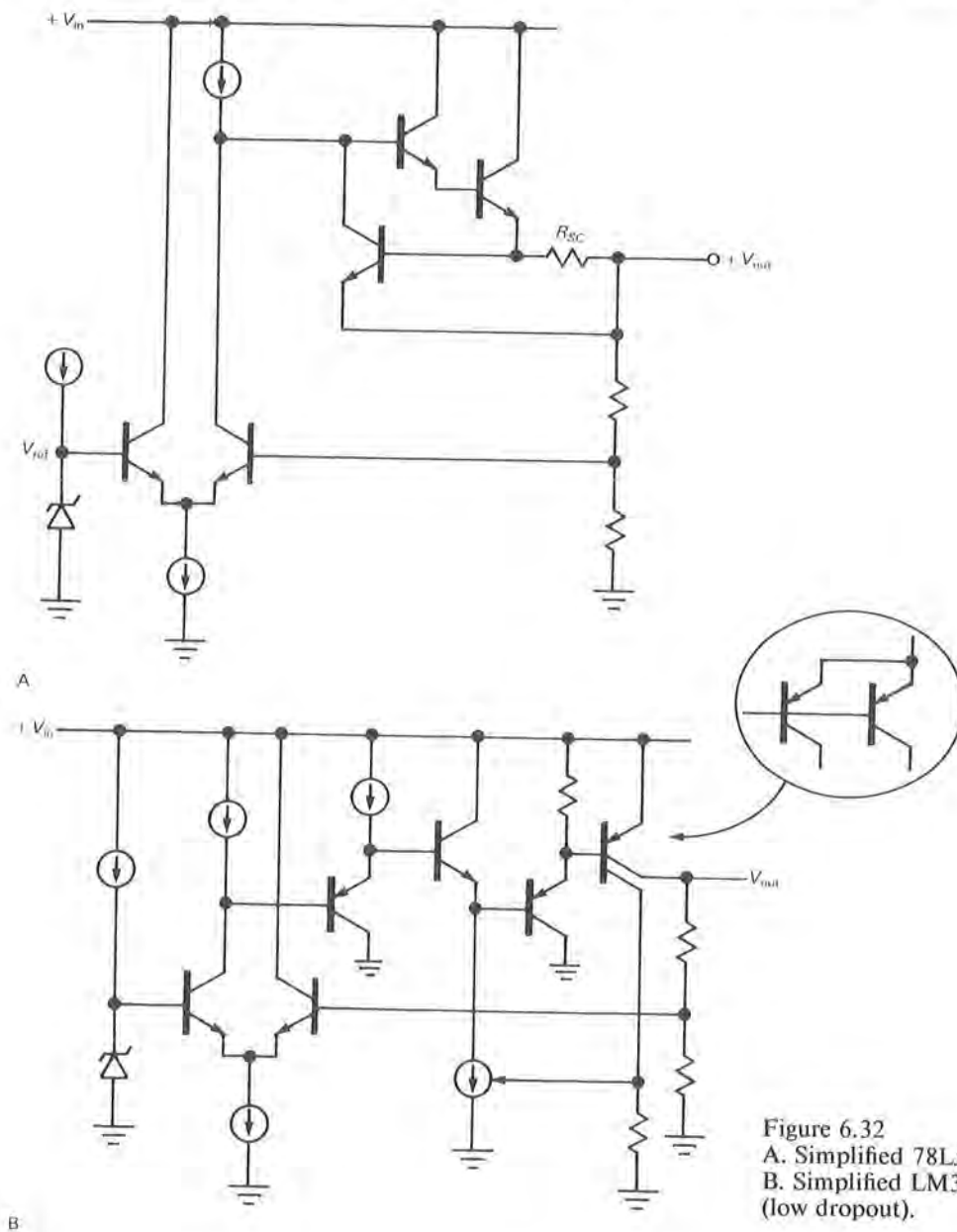


Figure 6.32
A. Simplified 78Lxx.
B. Simplified LM330
(low dropout).

(150mA) regulator. The output can be brought within a saturation voltage of the unregulated input voltage by the *pnp* pass transistor. Having thus eliminated the Darlington V_{BE} drops of the *nnp* regulator circuit, the designers weren't about to waste a diode drop with the usual (series resistor) short-circuit protection scheme. So they used a clever trick, deriving a sample of the output current via a second collector: That current is a fixed fraction of the output current and is used to shut off base drive as shown. This current-limit scheme is not particularly precise (I_L is specified as 150mA min, 700mA max), but it's good enough to protect the regulator, which also has internal thermal limiting.

Low-dropout regulators are available in most of the popular types, for example 3-terminal fixed voltage [LM2931, LM330, LT1083/4/5 (5V and 12V), TL750], 3-terminal adjustable (LT1083/4/5, LM2931), and micropower (LP2950/1, MAX664, LT1020). Tables 6.8 and 6.9 include all low-dropout regulators available at the time of writing.

Processor-oriented regulators

Electronic devices that include microprocessors (the subject of Chapters 10 and 11) require more than a simple regulated voltage. In order to retain the contents of volatile memory (and in order to keep track of elapsed time) they need a separate source of low-current dc when the regular power source is off; this may happen because the device is shut off, or because of a power failure. They also need to know when ordinary power has resumed, so they can "wake up" in a known state. Furthermore, a microprocessor-based device may need to have a few milliseconds warning that normal power is about to fail, so it can put data into safe memory.

Until recently you had to design extra circuitry to do these things. Now life is easy – you can get "(micro)processor-

oriented" regulator ICs, with various combinations of these functions built in. These ICs sometimes go by the name of "power supply supervisory chips" or "watchdog" chips. An example is the LM2984, which has two high-current +5 volt outputs (one for the microprocessor, one for other circuitry) and a low-current +5 volt output (for memory), a delayed RESET flag output to initialize your microprocessor when power resumes, and an ON/OFF controlling input for the high-current outputs. It also has an input that monitors microprocessor activity, resetting the processor if it grinds to a halt. An example of a watchdog chip *without* regulator is the MAX691 from Maxim, which monitors the regulated supply voltage and microprocessor activity, and provides reset (and "interrupt") signals to the microprocessor, just like the LM2984. However, it adds both power-fail *warning* and battery switchover circuitry to the other capabilities of the LM2984. Used with an ordinary +5 volt regulator, the MAX691 does everything you need to keep a microprocessor happy. We'll learn much more about the care and feeding of microprocessors in Chapters 10 and 11.

□ Micropower regulators

As we suggested earlier, most regulator chips draw a few milliamps of quiescent current to run the internal voltage reference and error amplifiers. That's no problem for an instrument powered from the ac mains, but it's undesirable in a battery-operated instrument, powered by a 400mA-hour 9 volt alkaline battery, and it's intolerable in a micropower instrument that must run a thousand hours, say, on a battery.

The solution is a micropower regulator. The most miserly of these are the ICL7663/4, positive and negative adjustable regulators with quiescent currents of $4\mu\text{A}$. At that current a 9 volt battery

would last 100,000 hours (more than 10 years), which exceeds the “shelf life” (self-discharge time) of all batteries except some lithium-based types. Micropower design is challenging and fun, and we’ll tell you all about it in Chapter 14.

Dual-polarity regulated supplies

Most of our op-amp circuits in Chapter 4 ran from symmetrical bipolarity supplies, typically ± 15 volts. That’s a common requirement in analog circuit design, where

you often deal with signals near ground, and the simplest way to generate symmetrical split supplies is to use a pair of 3-terminal regulators. For example, to generate regulated ± 15 volts, you could use a 7815 and a 7915, as in Figure 6.33A. We tend to favor the use of adjustable 3-terminal regulators, because (a) you only need to stock one type for each polarity and current range, and (b) you can trim the voltage exactly, if needed; Figure 6.33B shows how the circuit looks with a 317 and 337.

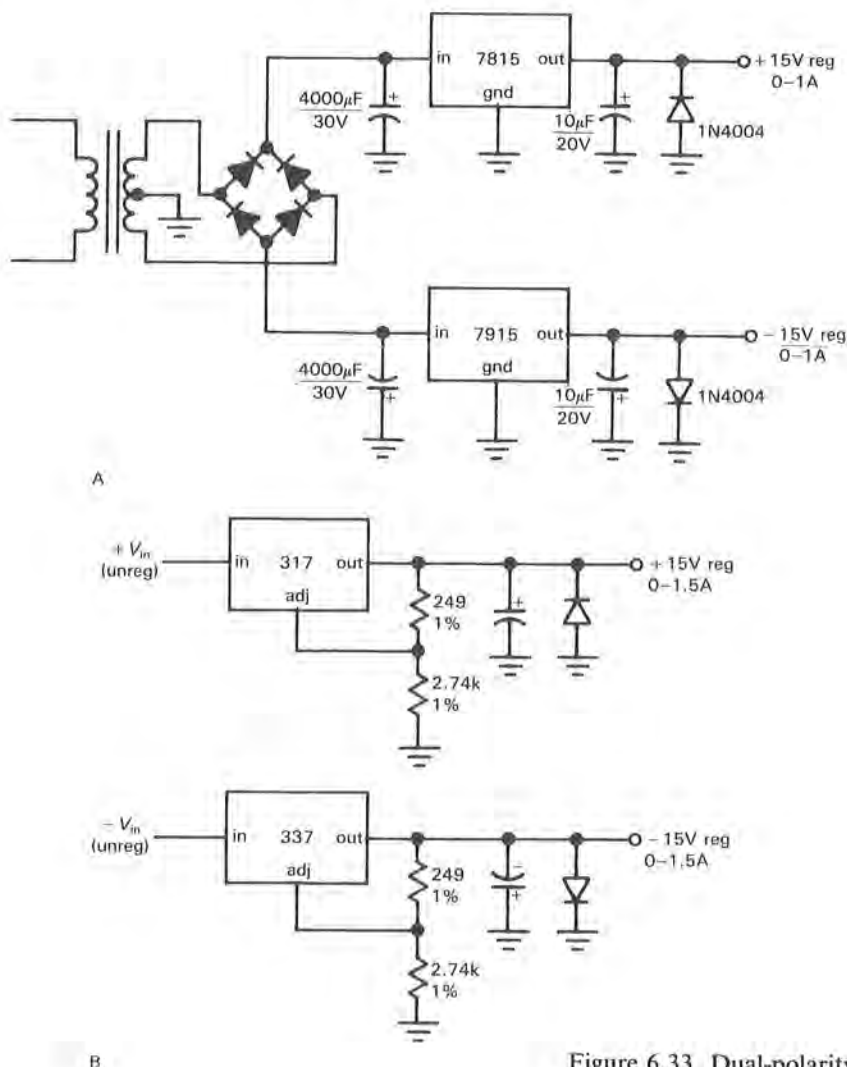


Figure 6.33. Dual-polarity regulated supplies.

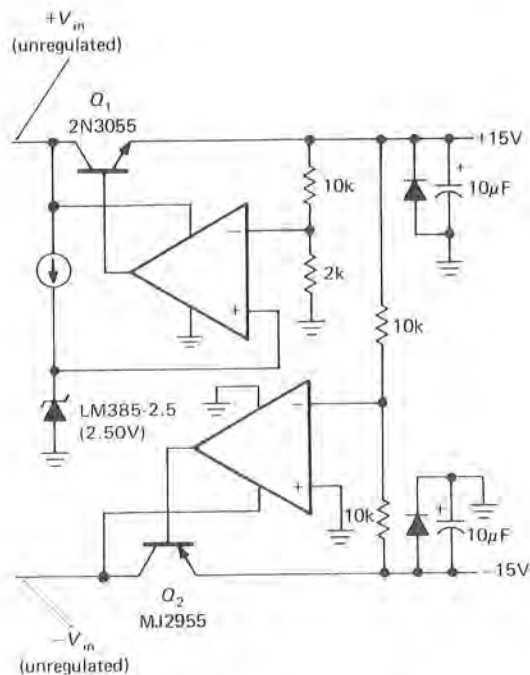


Figure 6.34. Dual-tracking regulator.

- **Dual tracking regulators.** Given the need for regulated split supplies, you might wonder why there aren't "dual 3-terminal regulators." Wonder no more – they exist and are known as "dual tracking regulators." To understand why they have such a complicated name, take a look at Figure 6.34, which shows the classic dual tracking regulator circuit. Q_1 is the pass transistor for a conventional positive regulated supply. The positive regulated output is then simply used as the reference for a negative supply. The lower error amplifier controls the negative output by comparing the average of the two output voltages with ground, thus giving equal 15 volt positive and negative regulated outputs. The positive supply can be any of the configurations we have already talked about; if it is an adjustable regulator, the negative output follows any changes in the positive regulated output. In practice, it is wise to include current-limiting circuitry, not shown in Figure 6.34 for simplicity.

As with single-polarity regulators, dual-tracking regulators are available as complete integrated circuits in both fixed and adjustable versions, though in considerably less variety. Table 6.10 lists the characteristics of most types now available. Typical are the 4194 and 4195 regulators from Raytheon, which are used as shown in Figure 6.35. The 4195 is factory-trimmed for ± 15 volt outputs, whereas the 4194's symmetrical outputs are adjustable via the single resistor R_1 . Both regulators are available in power packages as well as the small DIP packages, and both have internal thermal shutdown and current limiting. For higher output currents you can add outboard pass transistors (see below).

Many of the preceding regulators (e.g., the 4-terminal adjustable regulators) can be connected as dual-tracking regulators. The manufacturers' data sheets often give suggested circuit configurations. It is worth keeping in mind that the idea of referencing one supply's output to another supply can be used even if the two supplies are not of equal and opposite voltages. For instance, once you have a stable +15 volt supply, you can use it to generate a regulated +5 volt output, or even a regulated -12 volt output.

EXERCISE 6.6

Design a ± 12 volt regulator using the 4194.

Reverse-polarity protection. An additional caution with dual supplies: Almost any electronic circuit will be damaged extensively if the supply voltages are reversed. The only way that can happen with a single supply is if you connect the wires backward; sometimes you see a high-current rectifier connected across the circuit in the reverse direction to protect against this error. With circuits that use several supply voltages (a split supply, for instance), extensive damage can result if there is a component failure that shorts the two supplies together; a common

TABLE 6.10. DUAL-TRACKING REGULATORS

Type	Pkg	V _{out} (V)	Adj. output	Balance trim	Adj. cur. limit	Thermal limit	Max V ₊ -V ₋ Input (V)	Maximum output current ^a (each supply)					120Hz ripple reject typ (dB)	Temp stab ^e typ (mV)	Noise ^f (μV rms)
								@75°C case		P _{diss} (W)	Regulation typ				
								I _{out} (mA)	I _{out} (mA)		Load ^c (mV)	Line ^d (mV)			
Motorola															
MC1468L	DIP	±15	*	*	*	—	60	55	30	0.5	10 ^m	10 ^m	75	45	100
MC1468R	TO-66	±15	*	*	*	—	60	100	65	1.2	10 ^m	10 ^m	75	45	100
National															
LM325H ^g	TO-5	±15	—	—	*	*	60	100	50	0.5	6	2	75	45	150
LM325N ^g	DIP	±15	—	—	*	*	60	—	50	0.5	6	2	75	45	150
LM326H ^g	TO-5	adj	—	—	*	*	60	100	70	0.5	6	2	75	35	100
LM326N ^g	DIP	adj	—	—	*	*	60	—	70	0.5	6	2	75	35	100
Raytheon															
RC4194DB	DIP	adj	*	*	—	*	70	30 ⁱ	25 ⁱ	0.5	0.1%	0.2%	70	0.2%	250 ^j
RC4194TK	TO-66	adj	*	*	—	*	70	250 ⁱ	90 ⁱ	1.8	0.2%	0.2%	7	0.2%	250 ^j
RC4195NB	miniDIP	±15	—	—	*	*	60	—	20	0.35	2	2	75	75	60
RC4195TK	TO-66	±15	—	—	*	*	60	150	70	1.2	3	2	75	75	60
Silicon General															
SG3501AN	DIP	±15	—	*	—	*	60	60	30	0.6	30	20	75	150	50
SG3502N	DIP	adj	*	*	*	*	50	50 ⁱ	30	0.6	0.3%	0.2%	75	1%	50

(a) V_{in}=1.6V_{out} (each supply). (b) for 50°C ambient. (c) 10% to 50% I_{max}. (d) for ΔV_{in}=15V. (e) ΔV_{out} for 0°C to 100°C T_J. (f) 100Hz to 10kHz. (g) intended for use with a pair of external pass transistors. (h) θ_{JA}. (i) 10V drop (each supply). (j) 10Hz to 100kHz. (m) max.

(a) $V_{in} = 1.6V_{out}$ (each supply). (b) for 50°C ambient. (c) 10% to 50% I_{max} . (d) for $\Delta V_{in} = 15V$. (e) ΔV_{out} for 0°C to 100°C T_J . (f) 100Hz to 10kHz. (g) intended for use with a pair of external pass transistors. (h) Θ_{JA} . (i) 10V drop (each supply). (j) 10Hz to 100kHz. (m) max.

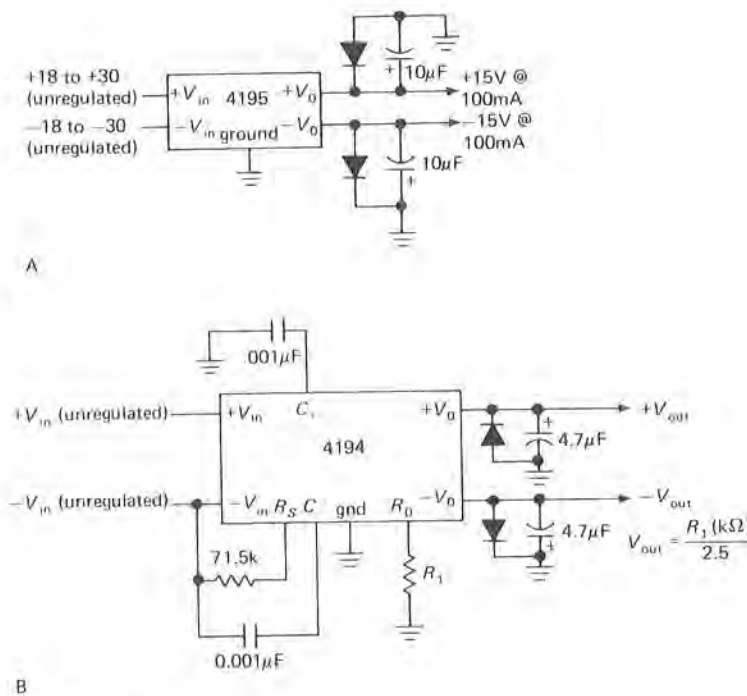


Figure 6.35

situation is a collector-to-emitter short in one transistor of a push-pull pair operating between the supplies. In that case the two supplies find themselves tied together, and one of the regulators will win out. The opposite supply voltage is then reversed in polarity, and the circuit starts to smoke. For this reason it is wise to connect a power rectifier (e.g., a 1N4004) in the reverse direction from each regulated output to ground, as we drew in Figure 6.33.

Outboard pass transistors

Three-terminal fixed regulators are available with 5 amps or more of output current, for example the adjustable 10 amp LM396. However, such high current operation may be undesirable, since the maximum chip operating temperature for these regulators is lower than for power transistors, mandating oversize heat sinks. Also, they are expensive. An alternative

solution is the use of external pass transistors, which can be added to the 3- and 4-terminal regulators (and dual-tracking regulators) just as with the classic 723. Figure 6.36 shows the basic circuit.

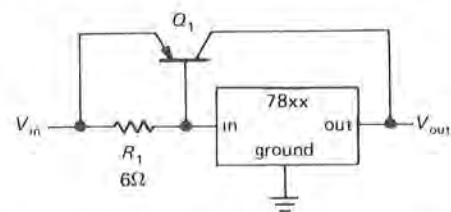


Figure 6.36. Three-terminal regulator with current-boosting outboard transistor.

The circuit works normally for load currents less than 100mA. For greater load currents, the drop across R_1 turns on Q_1 , limiting the actual current through the 3-terminal regulator to about 100mA. The 3-terminal regulator maintains the

output at the correct voltage, as usual, by reducing input current and hence drive to Q_1 if the output voltage rises, and vice versa. It never even realizes the load is drawing more than 100mA! With this circuit the input voltage must exceed the output voltage by the dropout voltage of the 78xx (2V) plus a V_{BE} drop.

In practice, the circuit must be modified to provide current limiting for Q_1 , which could otherwise supply an output current equal to h_{FE} times the regulator's internal current limit, i.e., 20 amps or more! That's enough to destroy Q_1 , as well as the unfortunate load that happens to be connected at the time. Figure 6.37 shows two methods of current limiting.

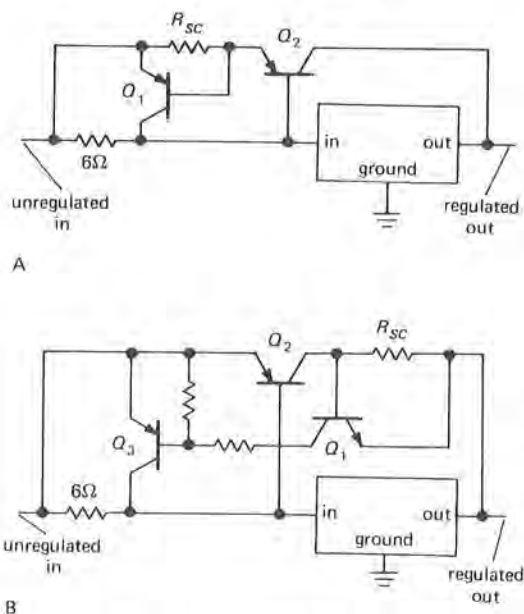


Figure 6.37. Current-limit circuits for outboard transistor booster.

In both circuits, Q_2 is the high-current pass transistor, and its emitter-to-base resistor has been chosen to turn it on at 100mA load current. In the first circuit, Q_1 senses the load current via the drop across R_{SC} , cutting off Q_2 's drive when the drop exceeds a diode drop. There are

a couple of drawbacks to this circuit: The input voltage must now exceed the regulated output voltage by the dropout voltage of the 3-terminal regulator plus two diode drops, for load currents near the current limit. Also, Q_1 must be capable of handling high currents (equal to the current limit of the regulator), and it is difficult to add foldback limiting because of the small resistor values required in Q_1 's base.

The second circuit helps solve these problems, at the expense of some additional complexity. With high-current regulators, a low dropout voltage is often important to reduce power dissipation to acceptable levels. To add foldback limiting to the latter circuit, just tie Q_1 's base to a divider from Q_2 's collector to ground, rather than directly to Q_2 's collector.

External pass transistors can be added to the adjustable 3- and 4-terminal regulators in exactly the same way. See the manufacturers' data sheets for further details.

Current source

A 3-terminal adjustable regulator makes a handy high-power constant-current source. Figure 6.38 shows one to source 1 amp. The addition of an op-amp follower, as in the second circuit, is necessary if the circuit is used to source small currents, since the "ADJ" (adjust) input contributes a current error of about $50\mu\text{A}$. As with the previous regulators, there is on-chip current limit, thermal-overload protection, and safe operating area protection.

EXERCISE 6.7

Design an adjustable current source for output currents from $10\mu\text{A}$ to 1mA using a 317. If $V_{in} = +15\text{V}$, what is the output compliance? Assume a dropout voltage of 2 volts.

Note that the current source in Figure 6.38A is a 2-terminal device. Thus, the load can be connected on either side. The figure shows how you might connect things

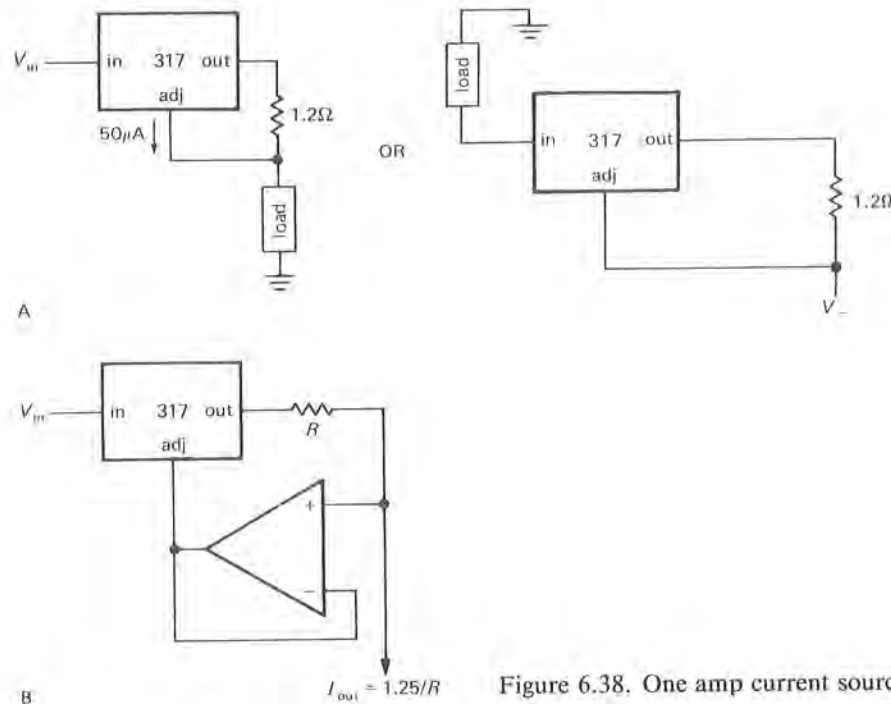


Figure 6.38. One amp current sources.

to *sink* current from a load returned to ground (of course, you could always use the negative-polarity 337, in the configuration analogous to Fig. 6.38A).

National makes a special 3-terminal device, the LM334, optimized for use as a low-power current source. It comes in the small plastic transistor package (TO-92), as well as the standard DIP IC package. You can use it all the way down to $1\mu\text{A}$, because the “adj” current is a small fraction of the total current. It has one peculiarity, however: The output current is temperature-dependent, in fact, precisely proportional to absolute temperature. So although it is not the world’s most stable current source, you can use it (Section 15.01) as a temperature sensor!

6.19 Switching regulators and dc-dc converters

All the voltage regulator circuits we have discussed so far work the same way: A

linear control element (the “pass transistor”) in series with the unregulated dc is used, with feedback, to maintain constant output voltage (or perhaps constant current). The output voltage is always lower in voltage than the unregulated input voltage, and some power is dissipated in the control element [the average value of $I_{out}(V_{in} - V_{out})$, to be precise]. A minor variation on this theme is the *shunt regulator*, in which the control element is tied from the output to ground, rather than in series with the load; the simple resistor-plus-zener is an example.

There is another way to generate a regulated dc voltage, fundamentally different from what we’ve seen so far; look at Figure 6.39. In such a *switching regulator* a transistor operated as a saturated switch periodically applies the full unregulated voltage across an inductor for short intervals. The inductor’s current builds up during each pulse, storing $\frac{1}{2}LI^2$ of energy in its magnetic field; the stored energy is transferred to a filter capacitor at the output,

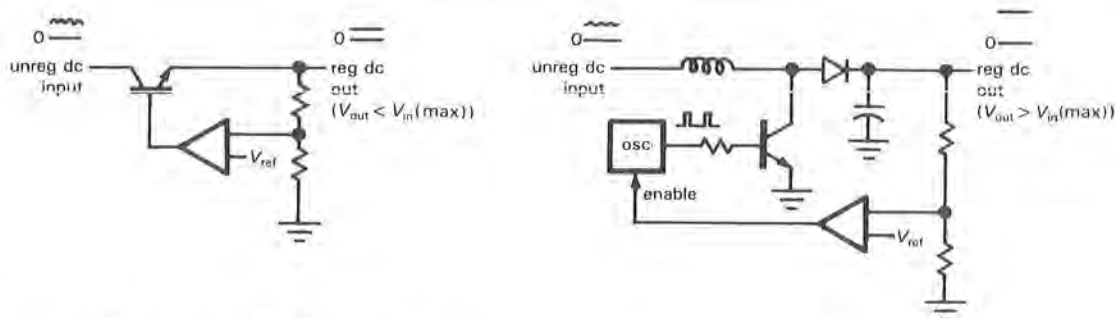


Figure 6.39. Two kinds of regulators.
A. Linear (series).
B. Step-up switcher.

which also smooths the output (to carry the output load between charging pulses). As with a linear regulator, feedback compares the output with a voltage reference – but in a switching regulator it controls the output by changing the oscillator’s pulse width or switching frequency, rather than by linearly controlling the base or gate drive.

Switching regulators have unusual properties that have made them very popular: Since the control element is either off or saturated, there is very little power dissipation; switching supplies are thus very efficient, even if there is a large drop from input to output. Switchers (slang for “switching power supplies”) can generate output voltages *higher* than the unregulated input, as in Figure 6.39B; they can just as easily generate outputs *opposite in polarity* to the input! Finally, switchers can be designed with no dc path from input to output; that means they can run directly from the rectified power line, with no ac power transformer! The result is a very small, lightweight, and efficient dc supply. For these reasons, switching supplies are used almost universally in computers.

Switching supplies have their problems, too. The dc output has some switching “noise,” and they can put hash back onto the power line. They used to have a bad reputation for reliability, with occasional

spectacular pyrotechnic displays during episodes of catastrophic failure. Most of these problems have been solved, however, and the switching supply is now firmly entrenched in electronic instruments and computers.

In this section we’ll tell you all about switching supplies, in two steps: First, we’ll describe the basic switching regulator, operating from a conventional unregulated dc supply. There are three circuits, used for (a) step-down (output voltage less than input), (b) step-up (output voltage greater than input), and (c) inverting (output polarity opposite to input). Then we’ll take a radical step, describing the heretical (and most widely used!) designs that run straight from the rectified ac power line, without an isolation transformer. Both kinds of power supplies are in wide use, so our treatment is practical (not just pedagogically pleasing). Finally, we’ll give you plenty of advice on the subject: When to use switchers, when to avoid them; when to design your own, when to buy them. With characteristic overstatement, we won’t leave you in any doubt!

□ Step-down regulator

Figure 6.40 shows the basic step-down (or “buck”) switching circuit, with feedback omitted for simplicity. When the MOS

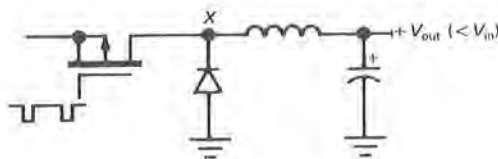


Figure 6.40. Step-down switcher.

switch is closed, $V_{out} - V_{in}$ is applied across the inductor, causing a linearly increasing current (recall $dI/dt = V/L$) to flow through the inductor. (This current flows to the load and capacitor, of course.) When the switch opens, inductor current continues to flow in the same direction (remember that inductors don't like to change their current suddenly, according to the last equation), with the "catch diode" now

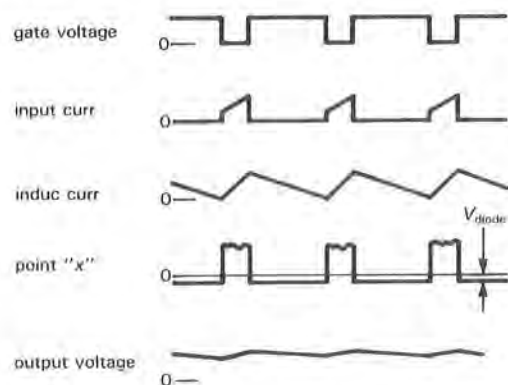


Figure 6.41

conducting to complete the circuit. The output capacitor acts as an energy "fly-wheel," smoothing the inevitable sawtooth ripple (the larger the capacitor, the less the ripple). The inductor current now finds fixed voltage $V_{out} - 0.6V$ across it, causing its current to decrease linearly. Figure 6.41 shows the corresponding voltage and current waveforms. To complete the circuit as a *regulator*, you would of course add feedback, controlling either the pulse width (at constant pulse repetition rate) or the

repetition rate (with constant pulse width) from an error amplifier that compares the output voltage with a reference.

Figure 6.42 shows a low-current +5 volt regulator using the MAX638 from Maxim. This nice chip gives you a choice of fixed +5 volt output (no external divider needed) or adjustable positive output, with external resistive divider. It includes nearly all components in a convenient mini-DIP package. In the MAX638 the oscillator runs at a constant 65kHz, with the error amplifier either permitting or cutting off gate drive pulses, according to the output voltage. The circuit shown gives about 85% efficiency, pretty much independent of the input voltage. Compare that with a linear regulator by doing the next problem:

EXERCISE 6.8

What is the maximum theoretical efficiency of a linear (series pass) regulator, when used to generate regulated +5 volt from a +12 volt unregulated input?

EXERCISE 6.9

What does a step-down regulator's high efficiency imply about the ratio of output current to input current? What is the corresponding ratio of currents, for a linear regulator?

□ Step-up regulator; inverting regulator

Apart from its high efficiency, the step-down switching regulator of the previous paragraph has no significant advantage (and some significant disadvantages – component count, switching noise) over a linear regulator. However, when there is a need for an output voltage greater than the unregulated input, or for an output voltage of opposite polarity to the unregulated input, switching supplies become very attractive indeed. Figure 6.43 shows the basic step-up (or "boosting") and inverting (sometimes called "flyback") circuits.

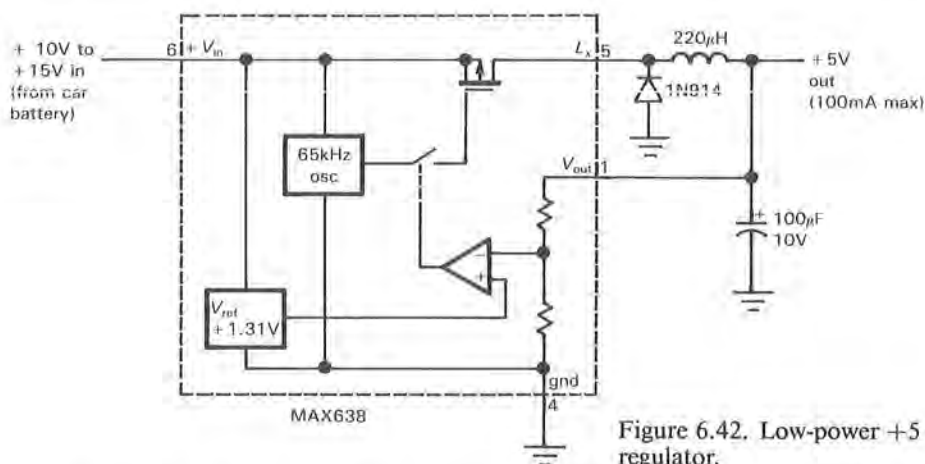


Figure 6.42. Low-power +5 volt switching regulator.

We showed the step-up circuit of Figure 6.39A earlier, in comparison with the linear regulator. The inductor current ramps up during switch conduction (point X near ground); when the switch is turned off, the voltage at point X rises rapidly as the inductor attempts to maintain constant current. The diode turns on, and the inductor dumps current into the capacitor. The output voltage can be much larger than the input voltage.

EXERCISE 6.10

Draw waveforms for the step-up switcher, showing voltage at point X , inductor current, and output voltage.

EXERCISE 6.11

Why can't the step-up circuit be used as a step-down regulator?

The inverting circuit is shown in Figure 6.43B. During switch conduction, a linearly increasing current flows from point X to ground. To maintain the current when the switch is open, the inductor pulls point X negative, as much as needed to maintain current flow. Now, however, the current is flowing *into* the inductor from the filter capacitor. The output is thus negative, and its average value can be larger or smaller in magnitude than the input (as determined by feedback); in other words,

the inverting regulator can be either step-up or step-down.

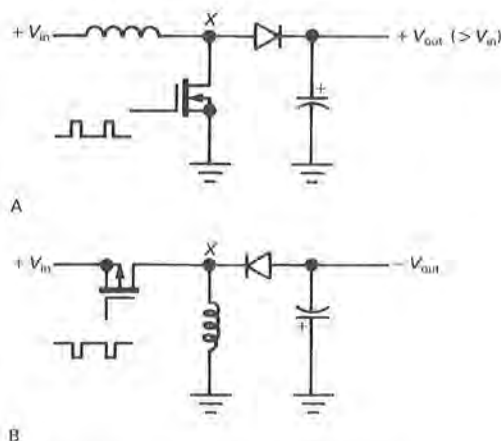


Figure 6.43. Two switching-element configurations.
A. Step-up ("boost").
B. Inverting.

EXERCISE 6.12

Draw waveforms for the inverting switcher, showing voltage at point X , inductor current, and output voltage.

Figure 6.44 shows how you might use low-power switching regulators to generate ± 15 volt op-amp supply voltages from a single +12 volt automotive battery, a trick that is impossible with linear regulators. Here we've again used low-power

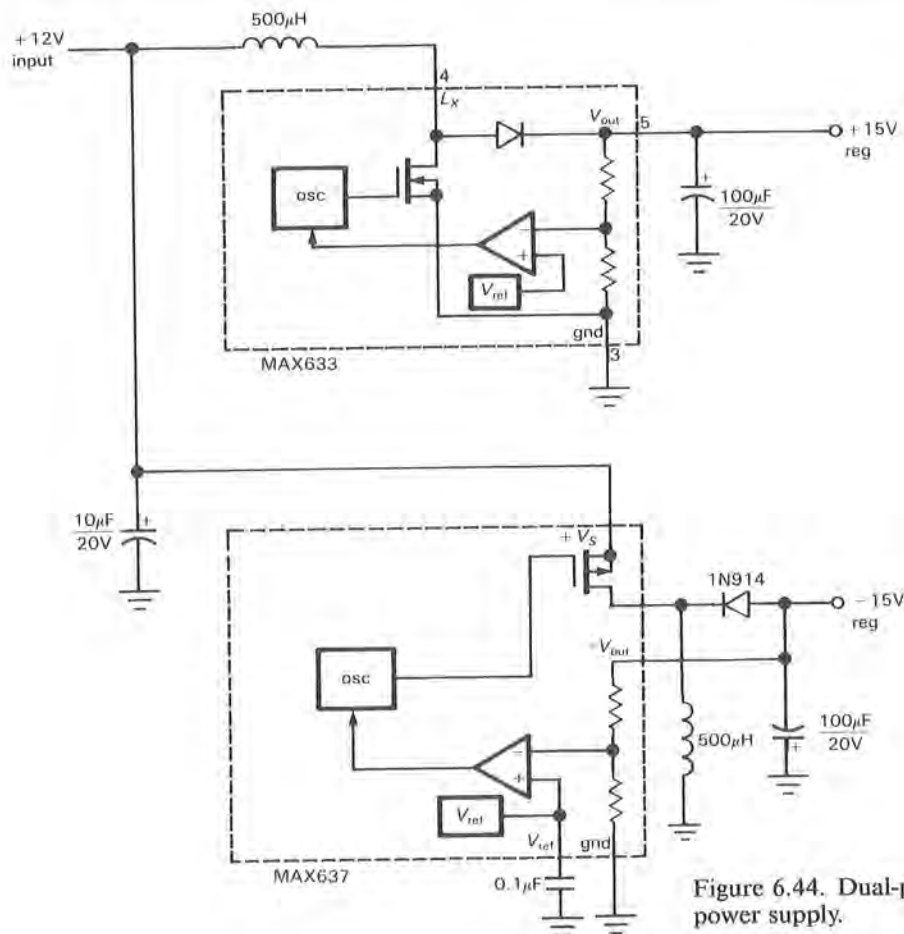


Figure 6.44. Dual-polarity switching power supply.

fixed-output ICs from Maxim, in this case the step-up MAX633 and the inverting MAX637. The external components shown were chosen according to the manufacturer's data sheets. They're not critical, but, as always in electronic design, there are trade-offs. For example, a larger value of inductor lowers the peak currents and increases the efficiency, at the expense of maximum available output current. This circuit is rather insensitive to input voltage, as long as it doesn't exceed the output voltage, and will work all the way down to +2 volts input, with greatly reduced maximum output current.

Before leaving the subject of inverting and step-up regulators, we should mention that there is one other way to accomplish the same goal, namely with "flying capacitors." The basic idea is to use MOS switches to (a) charge a capacitor from the dc input, and then (b) change the switches to connect the now-charged capacitor in series with another (step-up), or with reversed polarity across the output (inverting). Flying-capacitor voltage converters (e.g., the popular 7662) have some advantages (no inductors) and some disadvantages (low power, poor regulation, limited voltage). We'll discuss them later in the chapter.

General comments on switching regulators

As we've seen, the ability of switchers to generate stepped-up or inverted outputs makes them quite handy for making, say, low-current ± 12 volt supplies on an otherwise all-digital $+5$ volt circuit board. You'll often need such bipolarity supplies to power "serial ports" (more in Chapters 10 and 11) or linear circuitry using op-amps or A/D (analog-to-digital) and D/A (digital-to-analog) converters. Another good use for step-up switchers is to power displays that require relatively high voltage, for example using fluorescent or plasma technology. In these applications, where the dc input (typically $+5$ V) is already regulated, you often use the phrase "dc-to-dc converter," rather than "switching regulator," although it's really the same circuit. Finally, in battery-operated equipment you often want high efficiency over a wide range of battery voltage; for example, a 9 volt alkaline "transistor" battery begins life at about 9.5 volts, dropping steadily to about 6 volts at the end of its useful life. A $+5$ volt low-power step-down regulator maintains high efficiency, with current step-up over most of the battery's life.

Note that the inductor and capacitor in a switching regulator are not functioning as an LC filter. In the simple step-down regulator, that might seem to be true, but obviously a circuit that *inverts* a dc level is hardly a filter! The inductor is a lossless energy-storage device (stored energy = $\frac{1}{2}LI^2$), able to transform impedance in order to conserve energy. This is an accurate statement from a physicist's point of view, in which the magnetic field contains stored energy. We're more used to thinking of capacitors as energy storage devices (stored energy = $\frac{1}{2}CV^2$), which is their role in switching supplies, as in conventional series regulators.

A bit of nomenclature: You sometimes see the phrases "PWM switch-mode

regulator" and "current-mode regulator." They refer to the particular way in which the switching waveform is modified according to the feedback (error) signal. In particular, PWM means pulse-width modulation, in which the error signal is used to control the conduction pulse width (at fixed frequency), whereas in current-mode control the error voltage is used to control the peak inductor current (as sensed by a resistor) via width on a pulse-by-pulse basis. Current-mode regulators have some significant advantages and are becoming more popular now that good current-mode controller ICs have become available.

Keep in mind, when considering any switching supply, the noise generated by the switching process. This takes three forms, namely (a) output ripple, at the switching frequency, typically of order 10mV–100mV peak-to-peak, (b) ripple, again at the switching frequency, impressed onto the input supply, and (c) radiated noise, at the switching frequency and its harmonics, from switching currents in the inductor and leads. You can get into plenty of trouble with switching supplies in a circuit that has low-level signals (say 100 μ V or less). Although an aggressive job of shielding and filtering may solve such problems, you're probably better off with linear regulators from the outset.

Line-powered switching supplies

As we have seen, switching supplies have high efficiency even when the output voltage is nowhere near the input voltage. It may help our understanding to think of the inductor as an "impedance converter," since the average dc output current can be larger (step-down) or smaller (step-up) than the average dc input current. This is in stark contrast to linear series regulators, where the average values of input and output currents are always equal (ignoring the quiescent current of the regulator circuitry, of course).

This leads to a radical idea: We could eliminate the heavy 60Hz step-down transformer if we ran the regulator directly from rectified (unregulated) and filtered ac power. Two immediate comments: (a) The dc input voltage will be approximately 160 volts (for 115V ac power) – this is a dangerous circuit to tinker with! (b) The absence of a transformer means that the dc input is not isolated from the power line. Thus, the switching circuit itself must be modified to provide isolation.

The usual way to isolate the switching circuit is to wind a secondary onto the energy-storage inductor and use an isolation device (either transformer or optoisolator) to couple the feedback to the switching oscillator; see the simplified block diagram in Figure 6.45. Note that the oscillator circuitry is powered from the high-voltage unregulated dc, whereas the feedback control circuitry (error amplifier, reference) is powered from the regulated dc output. Sometimes an auxiliary low-current unregulated supply (with its own 60Hz low-voltage transformer) is used to power the control elements. The box labeled “isolation” is often a small pulse transformer, although optical isolation can also be used (more on this later).

It might seem as if the advantage of a transformerless unregulated supply is more than overcome by the need for at least *two* other transformers! Not so. The size of a transformer is determined by the core size, which decreases dramatically at high frequencies. As a result, line-powered switching supplies are much smaller and lighter than the equivalent linear supply; they also run cooler, owing to higher efficiency. For example, Power-One manufactures both kinds of supplies. Comparing their model F5-25 (5V, 25A) linear supply with their comparably priced SPL130-1005 (5V, 26A) switcher, we find that the switcher weighs 2.5 pounds, compared with 19 pounds for the linear, and occupies just one-fourth the volume. Furthermore, the switcher will run cool, while the 19-pound linear will run hot, dissipating up to 75 watts at full load.

□ Real-world switcher example

In order to give you a feel for the real complexity of line-powered switching supplies, we’ve reproduced in Figure 6.46 the complete schematic of a commercial switcher, in fact the power supply used by Tandy

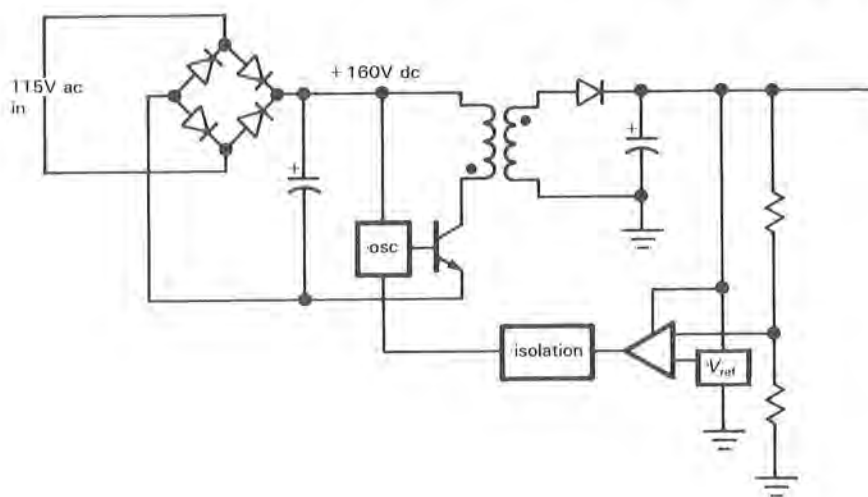


Figure 6.45. Direct ac-line-powered switching supply.

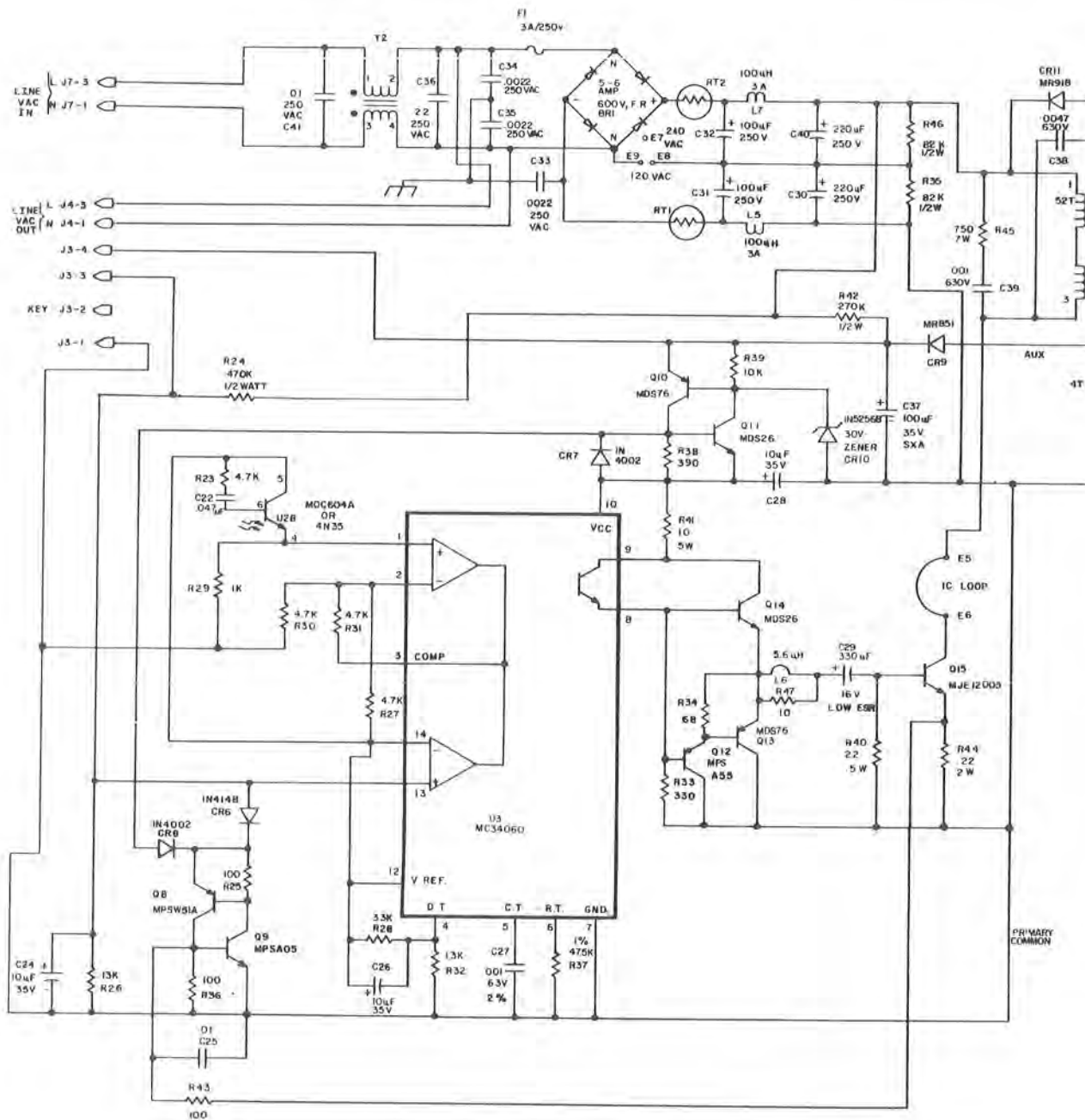


Figure 6.46. Switching power supply used in the Tandy model 2000 personal computer. Feedback from the +5 volt output is provided via opto-isolator U_{2A} - U_{2B} . (Courtesy of Tandy Corporation. Copyright 1984.)

(Radio Shack) to power their model 2000 personal computer. (We tried to get power-supply schematics from both IBM and Apple, but were ignored or haughtily rebuffed. Tandy, by comparison, publishes excellent documentation, with complete schematics and extensive circuit description.) It provides regulated outputs of +5 volts at 13 amps, +12 volts at 2.5 amps, and -12 volts at 0.2 amp (95W total), which are used to power the logic circuits and floppy-disk drives in the computer.

Let's take a walk through Figure 6.46 to see how a line-powered switcher copes with real-world problems. The circuit topology chosen by Tandy's designers is precisely that shown in Figure 6.45, though there are a few more components! Begin by comparing the figures: The line-powered bridge rectifier (BR_1) charges filter capacitors C_{30} , C_{31} , C_{32} , and C_{40} (T_2 is not a transformer – note the connections – but rather an interference filter). The charged capacitors are switched across the transformer primary (pins 1 and 3) by power transistor Q_{15} , whose switching waveform (a fixed-frequency square wave of variable pulse width) is provided by IC U_3 (a “PWM switch-mode regulator”). The secondary winding (there are actually three windings, one for each output voltage) is half-wave-rectified to generate the dc output: The +12 volts are produced by CR_2 from the 7-turn winding of pins 11 and 18, the -12 volts by CR_4 from the 5-turn winding of pins 13 and 20, and the +5 volts by the paralleled combination of CR_{13} and CR_{14} , each powered from its own (2-turn!) winding.

With multioutput switchers, only one output can be used for voltage-regulating feedback. It is conventional to use the +5 volts logic supply for this purpose, as Tandy has done here: R_{10} selects a fraction (nominally 50%) of the +5 volt output to compare with U_4 's internal +2.5 volt reference, turning on photodiode U_{2a} if the output is too high. This photodiode

couples optically to phototransistor U_{2b} , which varies the pulse width of U_3 to maintain +5 volts output. Thus the block labeled “isolation” in Figure 6.45 is an opto-coupler (see Section 9.10).

At this point we have accounted for perhaps 25% of the components in Figure 6.46. The rest are needed to cope with problems such as (a) short-circuit protection, (b) overvoltage and undervoltage shut-down, (c) auxiliary power for the regulating circuitry, (d) ac power filtering, and (e) linear post-regulation of the (tracking) ± 12 volt supplies. Let's explore the circuit in some more detail.

Beginning at the ac line input, we find four capacitors and a series inductor pair, together forming an RFI filter. It's always a good idea, of course, to clean up the ac power entering an instrument (see Section 6.11); here, however, the careful filtering is additionally needed to keep radiofrequency hash generated *inside* the machine (mostly from the switching action in the power supply) from radiating *out* through the power line. Note next the optional jumper at E_8E_9 , which converts the input from full-wave bridge (jumper open) to full-wave doubler (jumper shorted); manufacturers who wish to export their electronic wares must provide 110/220 volt compatibility, which is remarkably simple in the case of switching supplies.

Thermistors RT_1 and RT_2 are used to limit the high inrush current when the supply is first switched on, at which point the power line sees a few hundred microfarads of uncharged capacitor. Without the thermistors (or some other trick) the inrush current can easily exceed 100 amperes! The thermistors provide an ohm or two of series resistance, dropping to near zero when they warm up. Even with thermistors, the inrush current is impressive: The power supply has a specified “Input Surge Current” of 70 amps, maximum.

The 100 μ H series inductors L_5 and L_7 in the unregulated supply further clean

up transmitted switching hash, and the 82k shunt resistors (R_{35} and R_{46}) are “bleeders,” to make sure the power-supply filter capacitors discharge fully after power is turned off. Some additional passive “snubber” components (C_{38} , C_{39} , and R_{45}) are used to damp the large voltage spikes that otherwise might destroy the switching transistor Q_{15} . CR_{11} ’s function is more subtle – it cleverly returns unused transformer energy to filter capacitors C_{30} and C_{40} .

Moving down the page, we encounter some real trickery, namely the “auxiliary supply.” The circuits need some low-voltage, low-current dc to run the PWM controller chip and associated circuits. One possibility is to use a separate little linear supply, with its own line-powered transformer, etc. However, the temptation is overwhelming to hang another small winding (with half-wave rectifier) on T_1 , thus saving a separate transformer. That’s what the designer has done here, with a 4-turn winding (pins 9 and 10), rectified and filtered by CR_9 and C_{37} . This simple supply generates a nominal 15 volt output.

Sharp-eyed readers will have noticed a flaw in this scheme: The circuit cannot start itself, since the auxiliary power is only present if the supply is already running! This turns out to be an old problem: Designers of television sets love to play the same trick, deriving all their low-voltage supplies from auxiliary windings on the high-frequency horizontal drive transformer. The solution is the so-called kick-start circuit, in which some of the unregulated dc is brought over to start the circuit; once going, the supply keeps itself going from its transformer-derived dc power. In this circuit the kick-start comes via R_{42} , which begins charging up C_{37} at power-on. Nothing happens until the capacitor reaches a diode drop above CR_{10} ’s zener voltage, at which point the SCR-like combination of Q_{10} and Q_{11} is switched into conduction (figure out

how that works), dumping C_{37} ’s charge across C_{28} , thus momentarily powering the control circuitry (U_3 and all components to its left). Once the oscillation starts, CR_9 provides 15 volts with enough current to power the control circuitry continuously (which R_{42} cannot do).

Most of the components surrounding U_3 pander to its needs (C_{27} and R_{37} , for example, set the pulse repetition rate at 25kHz). At the input side, U_{2b} provides overall feedback to maintain the output at +5 volts, as described earlier. Q_8 and Q_9 are another SCR-like latch, this time triggered to shut down the oscillator (and the series latching switch $Q_{10}Q_{11}$) if driver Q_{15} ’s emitter current (sensed by R_{44}) is excessive, for example if the power supply sees a short-circuit load. The series combination $R_{43}C_{25}$ provides a 1 μ s time constant so that the circuit is not triggered by switching spikes. The shut-down circuitry also derives an input from divider $R_{26}R_{24}$, quenching oscillation if the ac input drops below 90 volts ac. At the output side of the controller U_3 , Q_{12} – Q_{14} provide high-current push-pull drive to Q_{15} ’s base from the single-ended on-chip *n*pn driver transistor (figure out how). Note the “ I_C loop,” an accessible length of wire in Q_{15} ’s collector, which lets you observe the current waveform on a scope by using a clip-on current probe (see, for example, the Tektronix catalog).

Things are considerably simpler on the output side of T_1 . The +5 volt supply uses paralleled power Schottky diodes (CR_{13} and CR_{14}) for fast recovery and low forward drop (the MBR3035PT is rated at 30A average current with 20kHz drive, 35V reverse breakdown, and 0.5V typical forward drop at 10A), with “snubber networks” (10 Ω /0.01 μ F) to protect the diodes from high-voltage spikes. The “ π -section” filter consists of 8800 μ F of input capacitance, a 3.5 μ H series inductor, and a 2200 μ F output capacitor. (The lower-current $\pm$ 12V outputs also use half-wave

Schottky rectifiers and π -section filters, with smaller-valued components.) This degree of filtering might seem extreme by linear regulator standards, but remember that there is no post-regulation – what comes out of the filter is the “regulated dc” – therefore lots of filtering is needed to reduce ripple, predominantly at the switching frequency, to the requisite $\approx 50\text{mV}$ or so at the output.

The +5 volt output is sensed via divider $R_3R_{10}R_{11}$, driving TI’s TL431 “3-terminal zener” (U_4), which, in combination with a few resistors and capacitors for feedback compensation, provides isolated feedback via opto-coupler U_{2ab} . The +5 volt output is also sensed, via $R_{18}R_{19}$, to trigger the overvoltage-sensor IC (U_1 ; $V_{\text{thresh}} = +2.5\text{V}$); the latter drives the gate of SCR Q_6 , which crowbars the +12 volt supply, shutting things down via current limiting in the primary side, as described earlier. U_1 is also wired to sense an under-voltage condition, via its dedicated auxiliary power from CR_5 and C_{19} ; the under-voltage signal (a saturated *npn* transistor to ground) is sent to the microprocessor, alerting the system to imminent power failure so that the program can be brought to an orderly shut-down during the few remaining milliseconds without loss of data.

The power-supply designers used a bit of trickery to improve regulation in the ± 12 volt supplies, which otherwise ride virtually open-loop on what is basically a +5 volt supply. For the +12 volt supply they used the +5 volt output as a reference for error amplifier Q_2 , which controls a “magnetic amplifier.” The latter consists of series saturable reactor L_3 , provided with an opposing “reset current” via Q_1 . The reset current determines how many volt-seconds the inductor will block before reaching the state of magnetic saturation, in which it acts as a good conductor. A magnetic amplifier deserves its name, because a small control current modifies a large output current. Mag-amp controllers

are available as complete integrated circuits, for example the UC3838 from Unitrode.

For the lower-current –12 volt supply the designers opted for the simpler solution of a linear 7912-type post-regulator, complete with diodes for protection against reverse polarity. Throughout, the designers have used bypass capacitors and bleeders on the dc outputs.

This power-supply circuit illustrates most of the details that seldom get mentioned in textbooks, but are essential in the real world. The extra component count in this circuit pays handsomely in ensuring a power supply that is robust under field conditions. Although this extra care in design might appear to be a display of unnecessary compulsiveness, in fact it is hard-nosed cost-effectiveness – each field failure under warranty costs the manufacturer at least a hundred dollars in real shipping and repair costs, not to mention the tarnished reputation produced by persistent failure.

General comments on line-powered switching power supplies

1. Line-powered switchers (also called “off-line” switchers, though we don’t like the term) make excellent *high-power* supplies. Their high efficiency keeps them cool, and the absence of a low-frequency transformer makes them considerably lighter and smaller than the equivalent linear supply. As a result, they are used almost exclusively to power computers, even desktop personal computers. They are finding their way into other portable instruments, too, even such noise-sensitive applications as oscilloscopes.

2. Switchers are noisy! Their outputs have tens of millivolts of switching ripple at their outputs, they put garbage onto the power line, and they can even scream audibly! One cure for output ripple, if that’s a problem, is to add an external high-current *LC* low-pass filter; alternatively,

you can add a low-dropout linear post-regulator. Some dc-dc converters include this feature, as well as complete shielding and extensive input filtering.

3. Switchers with multiple outputs are available and are popular in computer systems. However, the separate outputs are generated from additional windings on a common transformer. Typically, feedback is taken from the highest current output (usually the +5V output), which means that the other outputs are not particularly well regulated. There is usually a "cross-regulation" specification, which tells, for example, how much the +12 volt output, say, changes when you vary the load on the +5 volt output from 75% of full load to either 50% or 100% of full load; a typical cross-regulation specification is 5%. Some multiple-output switchers achieve excellent regulation by using linear post-regulators on the auxiliary outputs, but this is the exception. Check the specs!

4. Line-powered switchers may have a minimum load current requirement. If your load-current may drop below the minimum, you'll have to add some resistive loading; otherwise the output may soar or oscillate. For example, the +5 volt, 26 amp switching supply above has a minimum load current of 1.3 amps.

5. When working on a line-powered switcher, *watch out!* Many components are at line potential and can be lethal. You can't clip the ground of your scope probe to the circuit without catastrophic consequences.

6. When you first turn on the power, the ac line sees a large discharged electrolytic filter capacitor across it (through a diode bridge, of course). The resulting "inrush" current can be enormous; for our Power One switcher it's specified as 17 amps, maximum (compared with a full-load input current of 1.6A). Commercial switchers use various "soft-start" tricks to keep the inrush current within civilized bounds. One method is to put a negative-tempco

resistor (a low-resistance thermistor) in series with the input; another method is to actively switch out a small (10 Ω) series resistor a fraction of a second after the supply is turned on.

7. Switchers usually include overvoltage "shut-down" circuitry, analogous to our SCR crowbar circuits, in case something goes wrong. However, this circuit often is simply a zener sensing circuit at the output that shuts off the oscillator if the dc output exceeds the trip point. There are imaginable failure modes in which such a "crowbar" wouldn't crowbar anything. For maximum safety you may want to add an autonomous outboard SCR-type crowbar.

8. Switchers used to have a bad reputation for reliability, but recent designs seem much better. However, when they decide to blow out, they sometimes do it with great panache! We had one blow its guts out in a "catastrophic deconstruction," spewing black crud all over its innards and innocent electronic bystanders as well.

9. Line-powered switchers are definitely complex and tricky to design reliably. You need special inductors and transformers (and lots of them; Fig. 6.46). Our advice is to avoid the design phase entirely, by *buying* what you need! After all, why build what you can buy?

10. A switching supply presents a peculiar load to the power line that drives it. In particular, an increase in line voltage results in a *decrease* in average current, because the switcher operates at roughly constant efficiency: That's a negative resistance load (averaged over the 60Hz wave), and it can cause some crazy effects. If there's a lot of inductance in the power line, the system may oscillate.

Advice

Luckily for you, we're not bashful about giving advice! Here it is:

1. For *digital* systems, you usually need +5 volts, often at high current (10A or more).

Advice: (a) Use a line-powered switcher. (b) Buy it (perhaps adding filtering, if needed).

2. For analog circuits with low-level signals (small-signal amplifiers, signals less than $100\mu\text{V}$, etc.). *Advice:* Use a linear regulator; switchers are too noisy – they will ruin your life. *Exception:* For some battery-operated circuits it may be better to use a low-power dc-dc switching converter.

3. For high-power anything. *Advice:* Use a line-powered switcher. It's smaller, lighter, and cooler.

4. For high-voltage, low-power applications (photomultiplier tubes, flash tubes, image intensifiers, plasma displays). *Advice:* Use a low-power step-up converter.

In general, low-power dc-dc converters are easy to design and require few components, thanks to handy chips like the Maxim series we saw earlier. Don't hesitate to build your own. By contrast, high-power switchers (generally line-powered) are complex and tricky and extremely trouble-prone. If you must design your own, be careful, and test your design very thoroughly. Better yet, swallow your pride and buy the best switcher you can find.

SPECIAL-PURPOSE POWER-SUPPLY CIRCUITS

□ 6.20 High-voltage regulators

Some special problems arise when you design linear regulators to deliver high voltages. Since ordinary transistors typically have breakdown voltages of less than 100 volts, supplies to deliver voltages higher than that require some clever circuit trickery. This section will present a collection of such techniques.

□ *Brute force: high-voltage components*

Power transistors, both bipolar and MOSFET, are available with breakdown

voltages to 1000 volts and higher, and they're not even very expensive. Motorola's MJ12005, for example, is an 8 amp *npn* power transistor with conventional (V_{CEO}) collector-to-emitter breakdown of 750 volts, and base back-biased breakdown (V_{CEX}) of 1500 volts; it costs less than 5 dollars in single quantities. Their MTP1N100 (similar to the European BUZ50) is a 1 amp *n*-channel power MOSFET with 1000 volt breakdown and a price tag of a few dollars. Power MOSFETs in particular are often excellent choices for high-voltage regulators, owing to their excellent safe operating area (absence of thermally induced second breakdown).

By running the error amplifier near ground (the output-voltage-sensing divider gives a low-voltage sample of the output), you can build a high-voltage regulator with only the pass transistor and its driver seeing high voltage. Figure 6.47 shows the idea, in this case a +100 to +500 volt regulated supply using NMOS pass transistor and driver. Q_2 is the series pass transistor, driven by inverting amplifier Q_1 . The op-amp serves as error amplifier, comparing an adjustable fraction of the output with a precision +5 volt reference. Q_3 provides current limiting by shutting off drive to Q_2 when the drop across the 33 ohm resistor equals a V_{BE} drop. The remaining components serve more subtle, but necessary, functions: The diode protects Q_2 from reverse gate breakdown if Q_1 decides to pull its drain down rapidly (while the output capacitor holds up Q_2 's source). The various small capacitors in the circuit provide compensation, which is needed because Q_1 is operated as an inverting amplifier with voltage gain, thus making the op-amp loop unstable (especially considering the circuit's capacitive load). This circuit is an exception to the general rule that transistor circuits do not present a shock hazard!

We can't resist an aside here: In slightly modified form (reference replaced by

Petitioner Intel Corp., Ex. 1037
IPR2023-00783

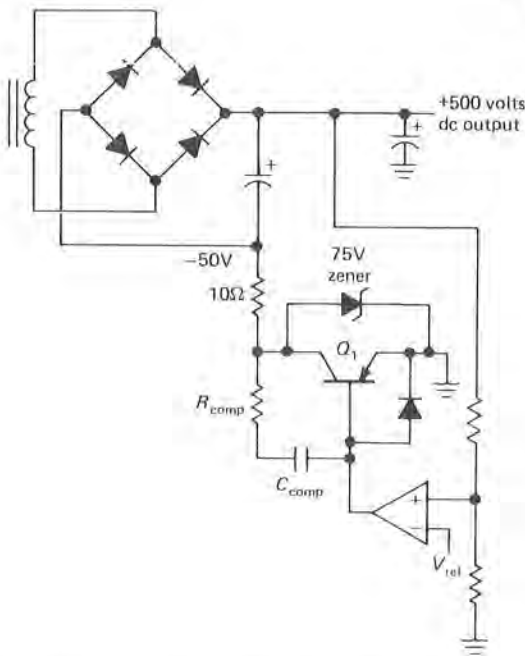


Figure 6.48. Regulating the ground return.

□ Lifting the regulator above ground

Another method sometimes used to extend the voltage range of regulators, including the simple 3-terminal type, is to raise the common terminal off ground with a zener (Fig. 6.49). In this circuit D_1 adds

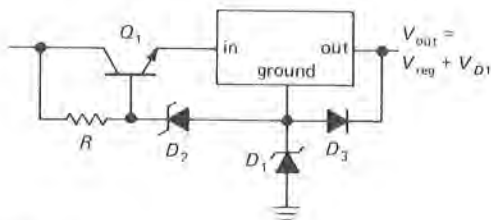


Figure 6.49

its voltage to the normal output of the regulator. D_2 sets the drop across the regulator via follower Q_1 and provides protection during short circuit because of D_3 .

□ Optically coupled transistor

There is another way to handle the problem of transistor breakdown ratings in high-voltage supplies, especially if the pass transistor can be a relatively low voltage unit because of fixed (known) output voltage. In such cases only the driver transistor has to withstand high voltage, and even that can be avoided by using optically coupled transistors. These devices, which we will talk about further in connection with digital interfacing in Chapter 9, actually consist of two units electrically isolated from each other: a light-emitting diode (LED), which lights up when current flows through it in the forward direction, and a phototransistor (or photo-Darlington) mounted in close proximity in an opaque package. Running current through the diode causes the transistor to conduct, just as if there were base current. As with an ordinary transistor, you apply collector voltage to put the phototransistor in the active region. In many cases no separate base lead is actually brought out. Optically coupled devices are typically insulated to withstand several thousand volts between input and output.

Figure 6.50 shows a couple of ways to use an optically coupled transistor in a high-voltage supply. In the first circuit, phototransistor Q_2 shuts off pass transistor Q_3 when the output rises too high. In the second version, for which only the pass-transistor circuitry is shown, phototransistor Q_2 increases the output voltage when driven, so the error-amplifier inputs should be reversed. Both circuits generate some output current through the pass-transistor biasing circuit, so some load from output to ground is needed to keep the output voltage from rising under no-load conditions. The output-sensing voltage divider can do the job, or a separate "bleeder" resistor can be connected across the output, which is always a good idea anyway in a high-voltage supply.

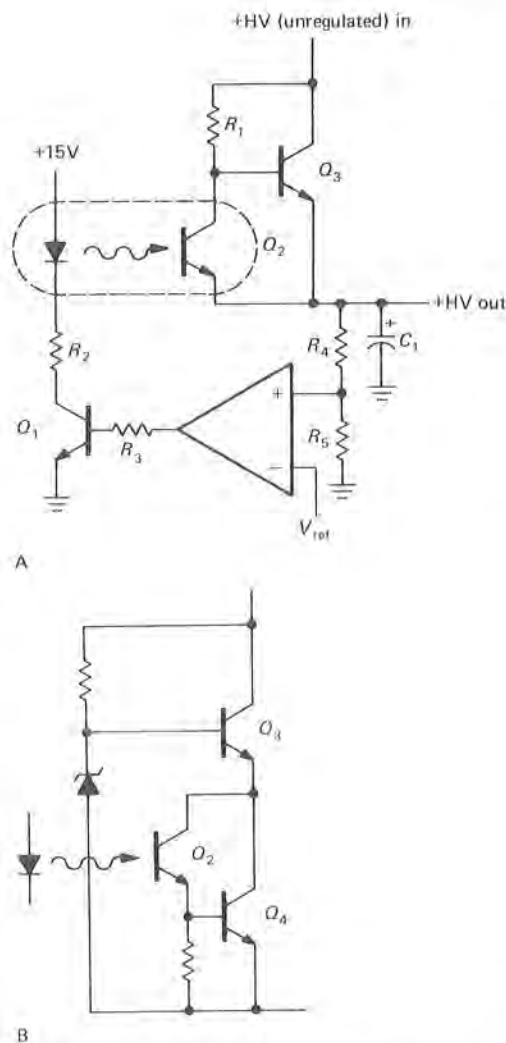


Figure 6.50. Opto-isolated high-voltage regulator.

□ Floating regulator

Another way to avoid applying large voltages to the control components of a high-voltage power supply is to “float” the control circuitry at the pass-transistor potential, comparing the drop across its own voltage reference with the drop down to ground. The excellent MC1466 regulator chip is intended for this kind of application, which normally requires an auxiliary low-current floating power supply

to provide dc (20–30V) for the chip itself. The output voltage is limited only by the pass transistors and the isolation (transformer insulation breakdown voltage) of the auxiliary supply. The MC1466 features very good regulation and precise current-limiting circuitry and is well suited for accurate “laboratory” power supplies. A warning, however: The MC1466, unlike more recent regulator designs, does not include on-chip thermal protection.

An elegant way to rig up a floating regulator is provided by the LM10 op-amp plus voltage-reference combination, a remarkable breakthrough in chip technology from the legendary Widlar (see Section 4.13) that will operate from a single 1.2 volt supply. Such a chip can be powered from the base-emitter drop of a Darlington pass transistor! Figure 6.51 shows an example. If you like analogies, think of a giraffe who measures his height by looking at the distance to the ground, then stabilizes it by craning his neck accordingly. The TL783 from Texas Instruments is a 125 volt IC regulator that works this way; for lower-current applications it replaces the discrete circuitry of Figure 6.51.

□ Transistors in series

Figure 6.52 shows a trick for connecting transistors in series to increase the breakdown voltage. Driver Q_1 drives series-connected transistors $Q_2 - Q_4$, which share the large voltage from Q_2 's collector to the output. The equal base resistors are chosen small enough to drive the transistors to full output current. The same circuit works with MOSFETs as well, but be sure to provide reverse-gate-protection diodes, as shown (you don't have to worry about *forward* gate breakdown, because the MOSFETs should turn on vigorously long before gate-channel breakdown). Note that the bias resistors produce some output current even when the transistors are cut off,

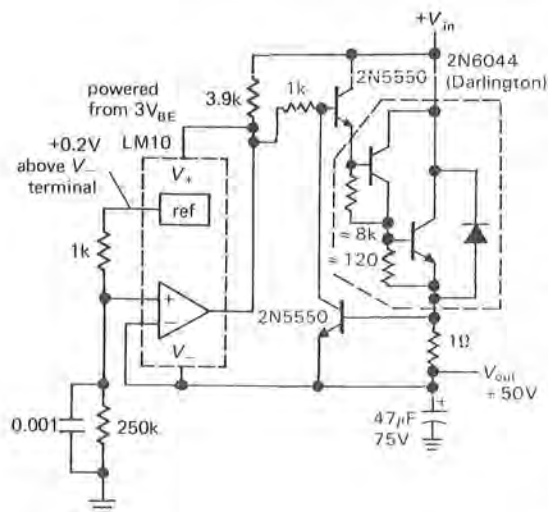


Figure 6.51. High-voltage floating regulator.

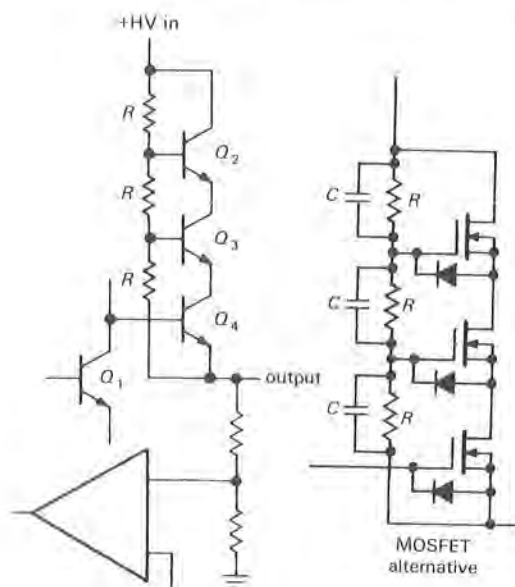


Figure 6.52. Connecting transistors in series to raise breakdown voltage.

so there must be a minimum load to ground to prevent the output from rising above its regulated voltage. It's often a good idea to parallel the divider resistors with small capacitors, as indicated, in order to maintain the divider action at high frequencies;

choose a capacitor value large enough to swamp differences in transistor input capacitance, which otherwise cause unequal division, reducing overall breakdown voltage.

Series-connected transistors can, of course, be used in circuits other than power supplies. You'll sometimes see them in high-voltage amplifiers, although the availability of high-voltage MOSFETs often makes it unnecessary to resort to the series connection at all.

In high voltage circuits like this, it's easy to overlook the fact that you may need to use 1 watt (or larger) resistors, rather than the standard $\frac{1}{4}$ watt type. A more subtle trap awaits the unwary, namely the maximum *voltage* rating of 250 volts for standard $\frac{1}{4}$ watt composition ("carbon") resistors, regardless of power dissipation. Carbon resistors run at higher voltages show astounding voltage coefficients, not to mention permanent changes of resistance. For example, in an actual measurement (Fig. 6.53) a 1000:1 divider (10Meg, 10k) produced a division ratio of 775:1 (29% error!) when driven with 1kV; note that the *power* was well within ratings. This non-ohmic effect is particularly important in the output-voltage-sensing divider of high-voltage supplies and amplifiers – beware! Companies like Victoreen make resistors in many styles designed for high-voltage applications like this.

□ Regulating the input

Another technique sometimes used in high-voltage supplies, particularly those intended for low currents, is to regulate the input rather than the output. This is usually done with high-frequency dc-to-dc switching supplies, since attempting to regulate the 60Hz ac input will result in poor regulation and plenty of residual ripple. Figure 6.54 shows the general idea. T_1 and associated circuitry generate unregulated dc at some manageable voltage, say 24

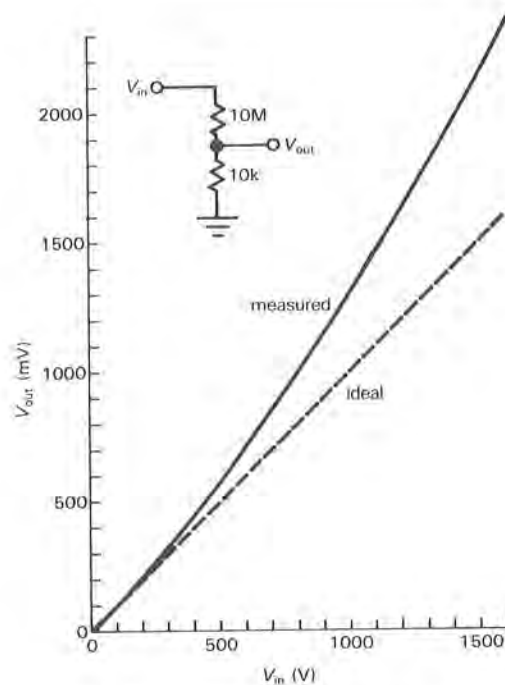


Figure 6.53. Carbon composition resistors exhibit a reduction in resistance above 250 volts.

volts; alternatively, batteries might provide the dc input. This powers a high-frequency square-wave power oscillator, with its output full-wave-rectified and filtered. This

filtered dc is the output, a sample of which is fed back to control the oscillator's duty cycle or amplitude in response to the output voltage. Since the oscillator runs at high frequency, the response is rapid, and its rectified waveform is easy to filter, especially since it is a full-wave-rectified square wave. T_2 must be designed for high-frequency operation, since ordinary laminated-core power transformers will have excessive core losses. Suitable transformers are built with iron powder, ferrite, or "tape-wound" toroidal cores and are much smaller and lighter than conventional power transformers of the same power rating. No high-voltage components are used, except, of course, for the output bridge rectifier and capacitor.

The astute reader may experience a sense of *déjà vu* while reading the last paragraph. In fact, it describes switching regulators (Section 6.19) in nearly all respects. The one significant difference is that switching supplies usually use inductors as energy-storage devices, whereas the input-regulated high-voltage supply uses T_2 as a "normal" (albeit high-frequency) transformer. In common with switching supplies, these high-voltage supplies display high-frequency ripple and noise.

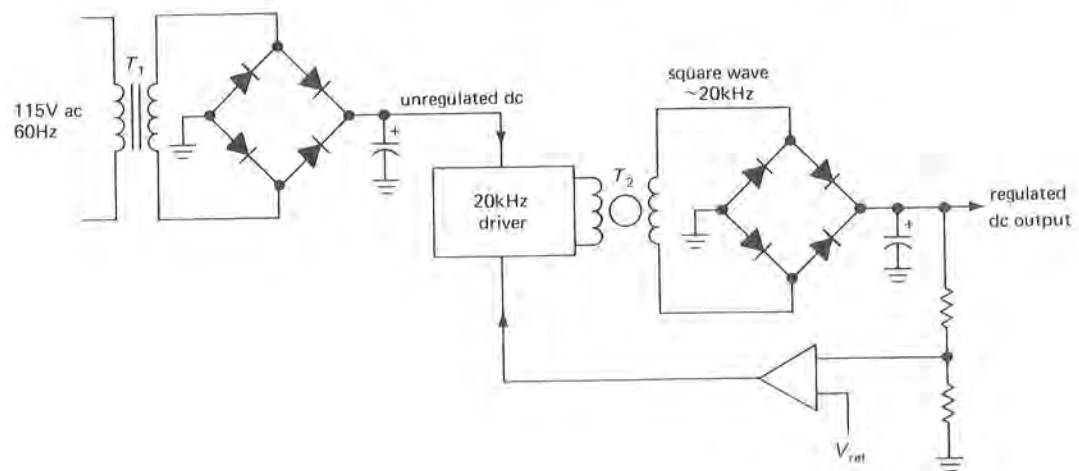
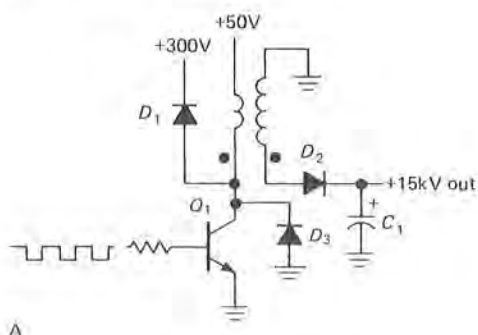


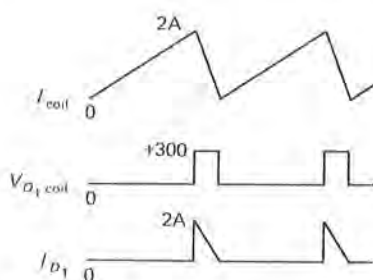
Figure 6.54. High-voltage switching supply.

Video flyback supplies

A variation on the conventional flyback switching regulator (Fig. 6.43A) is commonly used to generate the high dc voltages (10kV or more) needed in television and cathode-ray-tube (CRT) video displays. As we'll see, this circuit is especially clever, because it also generates the horizontal sweep signal used to drive the deflection coils.



A



B

Figure 6.55. Video flyback high-voltage supply.

The basic idea is to use a transformer with a large turns ratio, driving the primary with a saturated transistor, just like a conventional flyback circuit. The output is taken from the secondary, rectified to generate high-voltage dc; see Figure 6.55. Q_1 is driven by wide pulses, pulling the primary to ground. It may be self-excited or driven by an oscillator. D_1 is a “damper” diode that prevents Q_1 ’s collector from rising too high during the flyback. D_2 , connected to the high-voltage secondary

winding, rectifies the output, typically 10kV–20kV at a few microamps. The circuit is operated at frequencies of 15kHz or more, which means that filter capacitor C_1 can be as small as a few hundred picofarads (check this for yourself, by calculating the ripple).

Note that the collector-current waveform is a linearly rising ramp, which is often used to drive the magnetic deflection coils (called the “yoke”) of the CRT, thus producing the linear horizontal raster scan. In such cases the oscillator frequency sets the horizontal scan rate. A related circuit is the so-called blocking oscillator, which generates its own excitation pulses.

□ 6.21 Low-noise, low-drift supplies

The regulated supplies we have described thus far are pretty good – they typically have ripple and noise below a millivolt, and drift with temperature of 100ppm/°C or so. This is more than adequate for just about everything you will ever need to power. However, there are times when you may need better performance, and you can’t get it with any available regulator ICs. The solution is to design your own regulator circuit, using the best available IC references (in terms of stability and noise; see, e.g., the REF101KM in Table 6.7). This kind of stability (<1ppm/°C) is far better than the tempco of ordinary metal-film resistors (50ppm/°C), for example; so you must use great care to select op-amps and passive components whose errors and drifts do not degrade overall performance.

Figure 6.56 shows a complete design of an exceptional low-noise, low-drift dc regulated supply. It begins with the excellent REF10KM from Burr-Brown, which guarantees better than 1ppm/°C tempco, along with very low noise (6μV pp, 0.1–10Hz). Furthermore, it achieves this without thermostatic control, which helps keep the subsurface zener noise low. The reference

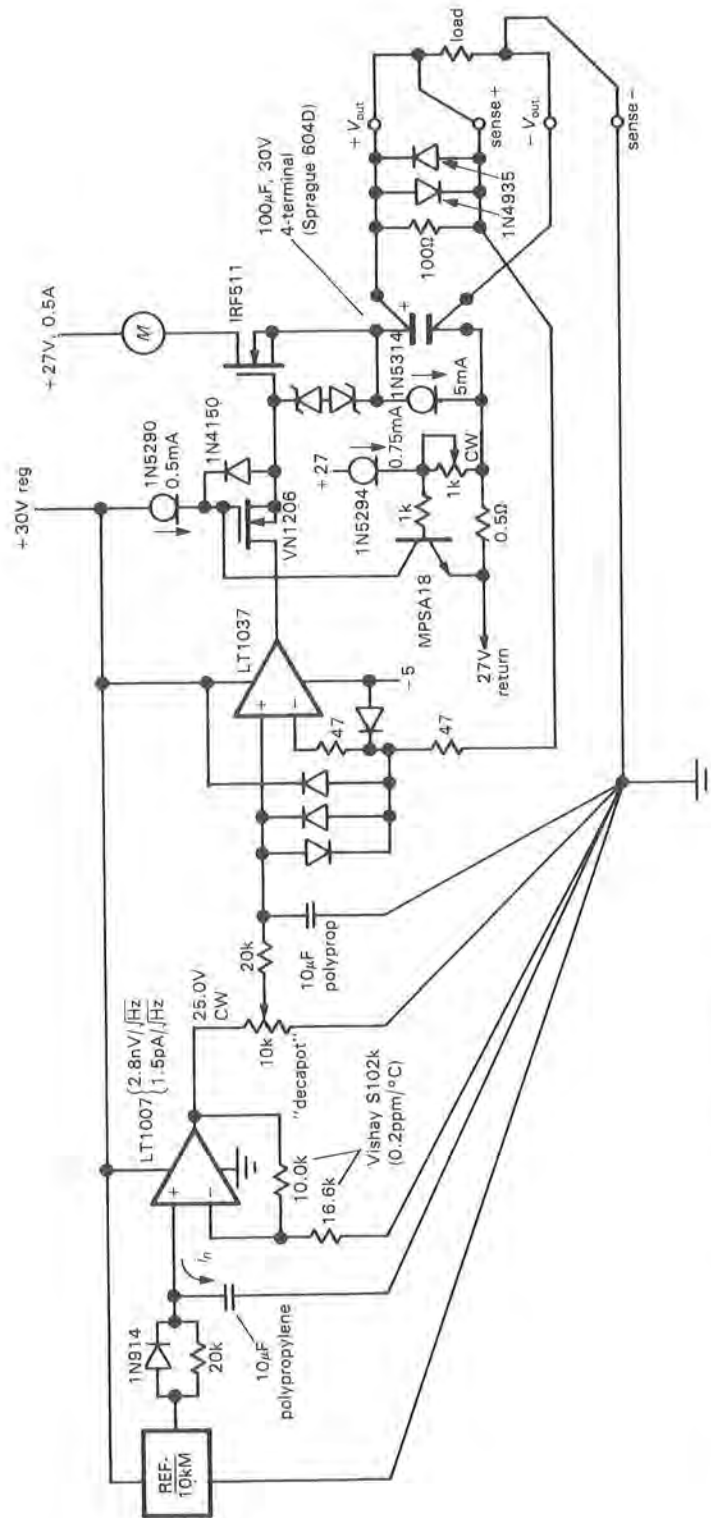


Figure 6.56. Ultrastable low-noise power supply.

is followed by a low-pass filter, to reduce the noise further. The large capacitor value is needed to suppress current noise from the op-amp: the value shown converts the current noise ($1.5\text{pA}/\sqrt{\text{Hz}}$ at 10Hz) to a voltage noise of $2.4\text{nV}/\sqrt{\text{Hz}}$, comparable with the op-amp's e_n . A polypropylene capacitor is used because the capacitor leakage (more precisely, *changes* in leakage over time and temperature) must be less than 0.1nA in order to avoid microvolt drifts in output voltage. The reference is boosted to $+25$ volts by the op-amp, whose feedback resistors have ultra-low tempco ($0.2\text{ppm}/^\circ\text{C}$, max); note the $+30$ volt supply voltage. The resultant $+25.0$ volt reference drives a voltage divider to produce the desired output voltage, which is then low-pass-filtered a second time, again using a low-leakage capacitor. Because a potentiometer is used to divide the reference voltage, resistor tempco isn't as critical here – it's a *ratio*metric measurement.

The rest of the circuit is simply a follower, using a precision low-noise error amplifier to compare the output voltage from a power MOSFET series pass transistor. A decompensated op-amp has been used, since the large output capacitor provides the dominant pole for compensation. Note the unusual current-limit circuit and the liberal use of constant-current “diodes” (really JFETs) to provide operating bias. Note also the use of “sense” wires to sample the voltage across the load. In a precision circuit like this it is important to pay careful attention to ground paths, since, for example, a 100mA load current flowing through 1 inch of #20 wire produces a voltage drop of $100\mu\text{V}$ – which is a 100ppm error for a 1 volt output! The circuit shown has excellent performance and surpasses the typical noise and drift figures given earlier by at least a factor of 100. According to EVI, Inc. (Columbia, MD), which kindly provided the circuit, it produces noise and hum below $1\mu\text{V}$, tempco

below $1\text{ppm}/^\circ\text{C}$, output impedance below $1\mu\Omega$, and drift below $1\text{ppm}/\text{working day}$.

We will talk more about such precision and low-noise design in the next chapter.

□ 6.22 Micropower regulators

As we've hinted earlier, it's possible to design battery operated circuits that use very low quiescent current, often as little as tens of microamps. That's what's needed, of course, to make the circuit run for months or years on a small battery, as it must if it is a wristwatch or calculator. For example, an alkaline 9 volt transistor battery is exhausted after supplying about 400mA-hours ; thus you can run a $50\mu\text{A}$ circuit with it for about a year (8800 hours). If such *micropower* circuits need regulated voltages, you clearly can't afford to squander the 3mA quiescent current of a 78L05, since that would degrade battery life to less than a week!

The solution is either to design a micropower regulator from discrete components or use one of the ICs intended for micropower applications. Luckily, some good ICs have come along in recent years. One of the best is the LP2950 series from National, available as a TO-92 (small transistor package) 3-terminal fixed 5 volt regulator (LP2950ACZ-5.0) or as a multiterminal adjustable 1.2–30 volt regulator (LP2951). Both versions have a quiescent current of $75\mu\text{A}$. For even lower quiescent currents there are the ICL7663/4 (or MAX663/4), adjustable regulators of both polarities with $4\mu\text{A}$ quiescent current. We will discuss micropower regulators, along with all aspects of battery-powered circuit design, in Chapter 14.

As an example of what you can do with discrete design, we show in Figure 6.57 a micropower circuit, designed for possible use in a lithium-battery-powered heart pacemaker, that converts an input voltage in the range $+5$ volts down to $+3$ volts (as

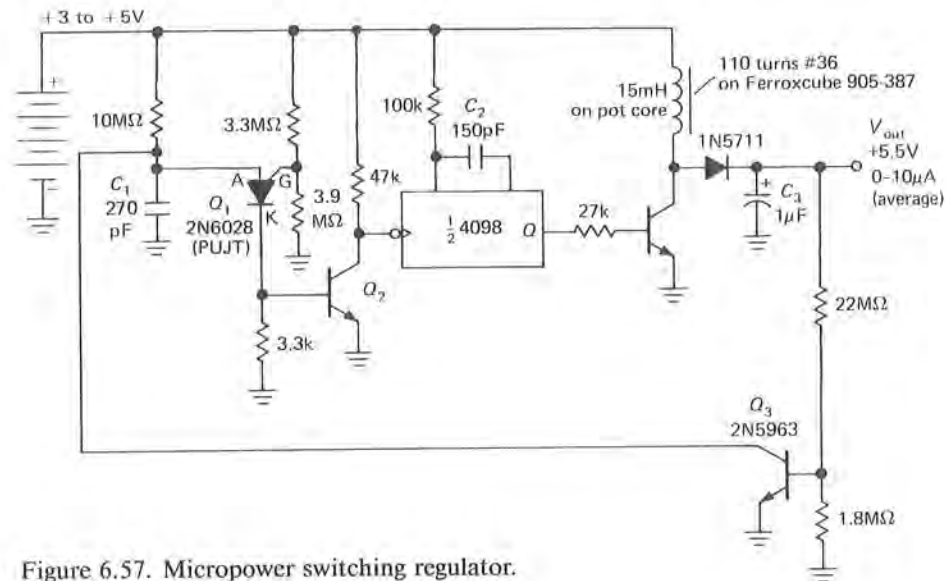


Figure 6.57. Micropower switching regulator.

the battery ages) to a regulated +5.5 volt supply. The power supply has a quiescent current of $1\mu\text{A}$ and provides line and load regulation of about 5%, with 85% conversion efficiency under full load for all battery voltages. As we remarked when discussing switching supplies, a conventional linear supply using an oscillator, doubler, and series pass regulator would be far less efficient because of regulator losses following the higher unregulated dc voltage. The flyback technique is effectively like a variable-ratio voltage multiplier, which yields extremely high efficiency, making it an attractive technique for micropower applications.

The 2N6028 programmable unijunction transistor (PUJT) is a versatile relaxation oscillator component. Its sense terminal (the anode) draws no current until its voltage exceeds the gate programming voltage by a diode drop, at which point it goes into heavy conduction from anode to cathode, discharging the capacitor. The resulting positive pulse at Q_2 's base pulls Q_2 's collector to ground, triggering the 4098, a device known as a "monostable

multivibrator" (see Section 8.20), which generates positive pulses of constant width at its output terminal labeled Q .

In this circuit, Q_3 senses the output voltage and robs charging current from C_1 , reducing the energy-transfer pulsing rate of the inductor as necessary to maintain the desired output voltage. Note the large resistor values throughout the circuit. Temperature compensation is not an issue here because the circuit operates in a stable 98.6°F mobile oven. (Warning: We remind the reader to look again at the “Legal notice” in the Preface.)

6.23 Flying-capacitor (charge pump) voltage converters

In Section 6.19 we discussed switching supplies, with their bizarre ability to produce a dc output voltage *larger* than the dc input, or even of opposite polarity. We mentioned there that flying-capacitor voltage converters let you do some of the same

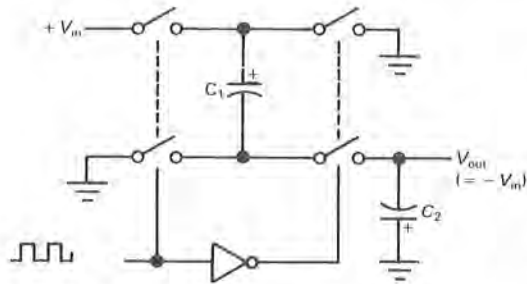


Figure 6.58. Flying-capacitor voltage inverter. C_1 and C_2 are external $10\mu\text{F}$ tantalum capacitors.

things. What is this strange “flying capacitor”?

Figure 6.58 shows a simplified circuit of the 7662 CMOS IC introduced by Intersil, and widely second-sourced. It has an internal oscillator and some CMOS switches, and it requires a pair of external capacitors to do its job. When the input pair of switches are closed (conducting), C_1 charges to V_{in} ; then, during the second half cycle, C_1 is disconnected from the input and connected, upside-down, across

the output. It thus transfers its charge to C_2 (and the load), producing an output of approximately $-V_{in}$. Alternatively, to use the 7662 to create an output of $2V_{in}$ you can arrange things so that C_1 charges as before, but then gets hooked in series with V_{in} during the second (transfer) half cycle.

This flying-capacitor technique is simple and efficient and requires few parts and no inductors. However, the output is not regulated, and it drops significantly under load currents greater than a few milliamps (Fig. 6.59). Also, like most CMOS devices, it has a limited supply voltage range; for the 7662, V_{in} can only range from 4.5 to 20 volts (1.5V to 10V for its predecessor, the 7660). Finally, unlike the inductive step-up or inverter circuits, which can generate any output voltage at all, the flying capacitor voltage converter can only generate discrete multiples of the input voltage. In spite of these drawbacks, flying-capacitor voltage converters can be very useful in some circumstances, for example to power a bipolarity op-amp or serial port (see Chapters 10 and 11) on a circuit board that has only +5 volts available.

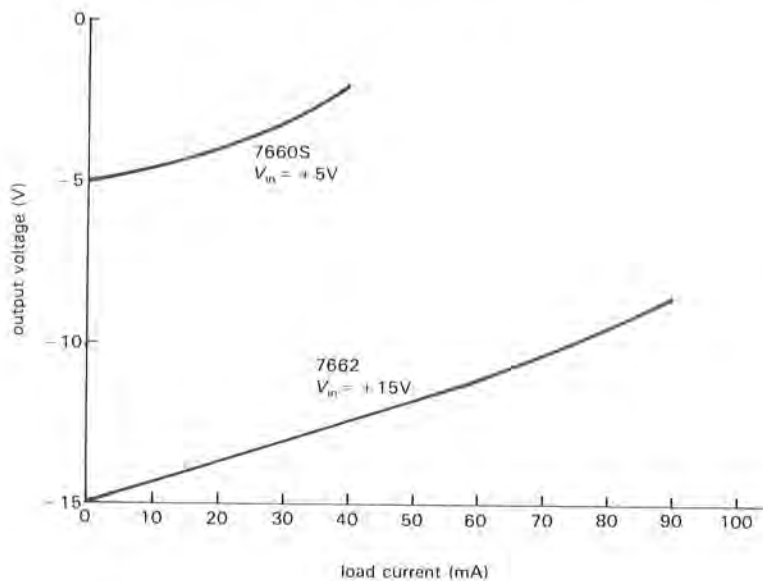


Figure 6.59. The output voltage of a flying-capacitor inverter drops significantly under load.

There are some other interesting flying-capacitor chips. The MAX680 from Maxim is a dual supply that generates ± 10 volts (up to 10mA) from +5 volts (Fig. 6.60). The similar LT1026 from LTC operates to ± 20 volts output (up to 20mA) and uses smaller capacitors ($1\mu\text{F}$ instead of $20\mu\text{F}$). The LT1054 from LTC combines a flying-capacitor converter with a linear regulator to provide a stiff regulated output up to 100mA (at lower efficiency, of course). The MAX232 series and the LT1080 combine a ± 10 volt switched-capacitor supply with an RS-232C digital serial port (see Chapter 11), eliminating the need for bipolarity supplies in many computer boards; some chips in the MAX232 series even have built-in capacitors. And the LTC1043 is an uncommitted flying-capacitor building block, which you can use to do all kinds of magic. For example, you can use a flying capacitor to transfer a voltage drop measured at an inconvenient potential (e.g., a current-sensing resistor at the positive supply voltage) down to ground, where you can easily use it. The LTC1043 data sheet has 8 pages of similar clever applications.

6.24 Constant-current supplies

In Sections 2.06 and 2.14 we described some methods for generating constant currents within a circuit, including voltage-programmed currents with floating or grounded loads and various forms of current mirrors. In Section 3.06 we showed how to use FETs to construct some simple current-source circuits, including “current-regulator diodes” (a JFET with gate tied to source) such as the IN5283 series. In Section 4.07 we showed how to get improved performance (at low frequencies, anyway) by using op-amps to construct current sources. And in Section 6.15 we mentioned the convenient LM334 3-terminal current source IC. There is often a need, however, for a flexible constant-current supply, which can supply substantial

voltage and current, as a complete instrument. In this section we will look at some of the more successful circuit techniques.

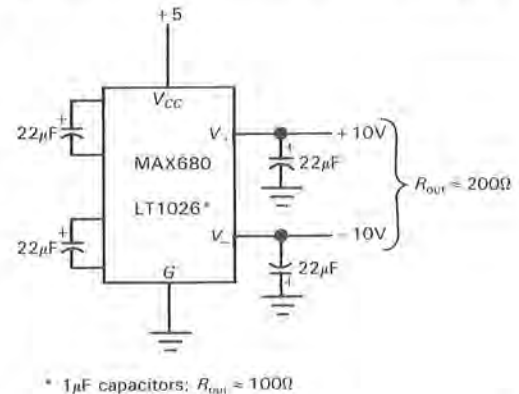


Figure 6.60. Flying-capacitor dual supply. The LT1026 is similar, but has $R_{\text{OUT}} \approx 100$ ohms and requires only $1\mu\text{F}$ capacitors.

□ Three-terminal regulator

In Section 6.18 we showed how you can use a 3-terminal adjustable regulator to make a delightfully simple current source. The 317-type regulator, for example, maintains a constant 1.25 volts (bandgap) between its output and its “ADJ” pin; by putting a resistor across these pins, you form a 2-terminal constant-current device (Fig. 6.38), which can be used as a sink or source. Performance degrades with less than about 3 volts across the circuit, since the regulator itself has a dropout voltage near 2 volts.

This type of current source is suitable for moderate to high currents: The LM317 has a maximum current of 1.5 amps and can operate with up to 37 volts drop. Its high-voltage cousin, the LM317HVK, can withstand 57 volts drop. Higher-current versions are available, e.g., the LM338 (5A) and LM396 (10A), although these have lower voltage ratings. Three-terminal regulators won’t work as current sources below about 10mA, the worst-case quiescent current. However, note that the

latter is not a source of current error, since it flows from input pin to output pin; the much smaller current that flows out of the ADJ pin ($50\mu\text{A}$, nominal) varies about 20% over the operating temperature range and is negligible by comparison.

In ancient times, before 3-terminal adjustable regulators were available, people sometimes used 5 volt *fixed* regulators (e.g., the 7805) as current sources in a similar arrangement (substituting “GND” for “ADJ”). This is an inferior circuit, because at low output currents the regulator’s quiescent current (8mA max) contributes a large error, and at high currents the 5 volt drop across the current-setting resistor results in unnecessary power dissipation.

□ Supply-line sensing

A simple technique that yields good performance involves constructing a conventional series pass regulator, with current sensing at the input to the pass transistor (Fig. 6.61). R_2 is the current-sensing

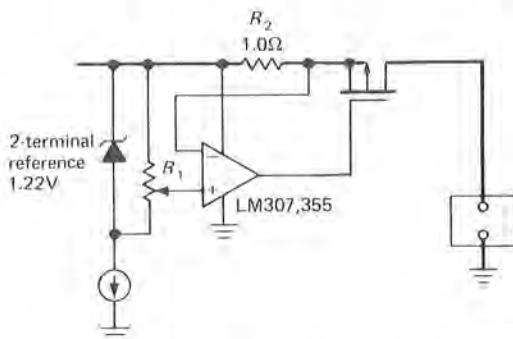


Figure 6.61. Input-rail current sensing.

resistor, preferably a low-temperature-coefficient type. For very high current or high-precision applications, you should use a 4-wire resistor, intended for current-sensing applications, in which the sensing leads are connected internally. The sensing voltage does not then depend on the connection resistance of the joints

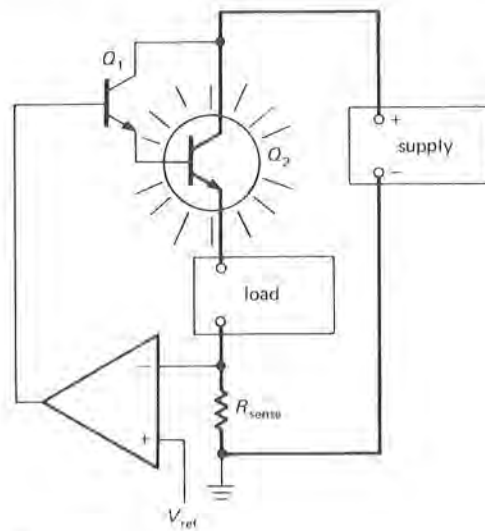
to the current-carrying leads, which for clarity are drawn with heavy lines in this schematic.

For this circuit you must use an op-amp that has an input common-mode range all the way to the positive supply (the 307, 355, and 441 have this virtue), unless, of course, you power the op-amp with a more positive auxiliary supply. The MOSFET in this circuit could be replaced by a *pnp* pass transistor; however, since the output current would then include the base current, you should use a Darlington connection to minimize that error. Note that an *n*-channel output transistor (connected as a follower) can be used instead of the *p*-channel shown, if the input connections to the op-amp are reversed. However, the current source will then have an undesirably low output impedance at frequencies approaching f_T of the op-amp loop, since the output is actually a source follower. This is a common error in current-source design, since the dc analysis shows correct performance.

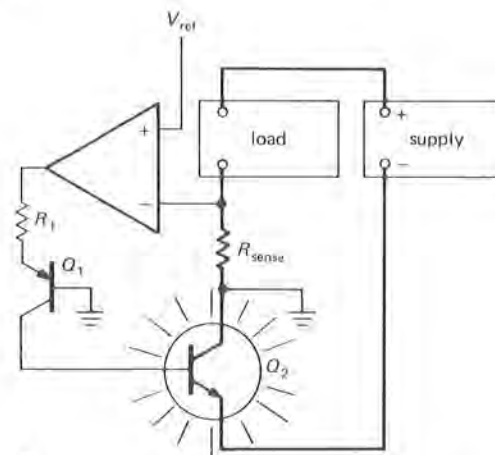
□ Return-line current sensing

A good way to make a precise current source is to sense the voltage across a precision resistor directly in series with the load, since this makes it easier to meet the simple criterion for eliminating current-source errors due to base drive currents; the base drive current must either pass through both the load and sense amplifier, or pass through neither. However, to meet this criterion it is necessary to “float” either the load or the power supply by at least the voltage drop across the current-sensing resistor. Figure 6.62 shows a couple of circuits that use floating loads.

The first circuit is a conventional series pass circuit, with the error signal derived from the drop across the small resistor in the load’s return path to ground. The high-current path is again drawn with bold lines. The Darlington connection is used here



A



B

Figure 6.62. Return-line current sensing.

not to avoid base-current error, since the actual current through the load is sensed, but rather to keep the drive current down to a few milliamps so that ordinary op-amps can be used for the error amplifier. The sensing resistor should be a precision power resistor of low temperature coefficient, preferably a 4-wire resistor. In the second circuit the regulating transistor Q_2 is in the ground return of the high-current supply. The advantage here is that its

collector is at ground, so you don't have to worry about insulating the transistor case from the heat sink.

In both circuits, R_{sense} will normally be chosen to drop a volt or so at typical operating currents; its value is a compromise between op-amp input offset errors, at one extreme, and a combination of reduced current-source compliance and increased dissipation at the other. If the circuit is meant to operate over large ranges of output current, R_{sense} should probably be a set of precision power resistors, with the appropriate resistor selected by a range switch.

□ Grounded load

If it is important for the load to be returned to circuit ground, a circuit with floating supply can be used. Figure 6.63 shows two examples. In the first circuit, the funny-looking op-amp represents an error amplifier with a high-current buffer output, run from a single split supply; it could be something as simple as a 723 (for currents up to 150mA) or one of the high-current op-amps listed in Table 4.4. The high-current supply has a common terminal that floats relative to circuit ground, and it is important that the error amplifier (or at least its buffer output) be powered from the floating supply so that base drive currents return through R_{sense} . An additional low-current supply with grounded common would be needed if other op-amps, etc., were in the same instrument. A negative reference (relative to circuit ground) programs the output current. Note the polarity at the error-amplifier inputs.

The second circuit illustrates the use of a second low-power supply when an ordinary low-current op-amp is used as error amplifier. Q_1 is the outboard pass transistor, which must be a Darlington (or MOSFET), since the base current returns through the load, but not through the sense resistor. The error amplifier is now

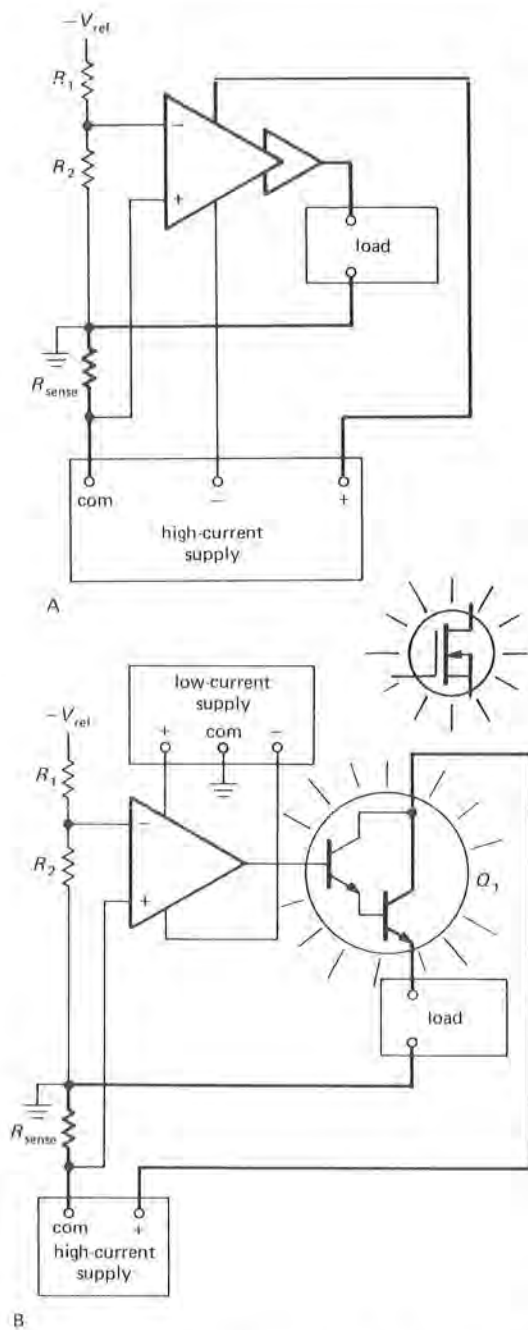


Figure 6.63. Current sources for grounded loads, employing floating high-current supplies.

powered from the same split supply with grounded common that powers the rest

of the instrument. This circuit is well suited as a simple bench-instrument current source, with the low-current split supply built in and the high-current supply connected externally. You would choose the latter's voltage and current capability to fit each application.

6.25 Commercial power-supply modules

Throughout the chapter we have described how to design your own regulated power supply, implicitly assuming that is the best thing to do. Only in the discussion of line-operated switching power supplies did we suggest that the better part of valor is to swallow your pride and buy a commercial power supply.

As the economic realities of life would have it, however, the best approach is often to use one of the many commercial power supplies sold by companies such as ACDC, Acopian, Computer Products Inc., Lambda, Power-One, and literally hundreds more. They offer both switching and linear supplies, and they come in four basic packages (Figure 6.64):

1. Modular "potted" supplies: These are low-power supplies, often dual (± 15) or triple ($+5$, ± 15), packaged in "potted" modules that are usually 2.5×3.5 ", and about 1" thick. The most common package has stiff wire leads on the bottom, so you can mount it directly on a circuit board; you can also screw it to a panel, or plug it into a socket. They are also available with terminal-strip screw connections along one side, for chassis mounting. A typical linear triple supply provides $+5$ volts at 0.5 amp and ± 15 volts at 0.1 amp and costs about \$100 in small quantity. Linear modular supplies fall in the 1–10 watt range, switchers in the 15–25 watt range.

2. "Open-frame" supplies: These consist of a sheet-metal chassis, with circuit board, transformer, and power transistors mounted in full view. They're meant to go



Figure 6.64. Commercial power supplies come in a variety of shapes and sizes, including potted modules, open-frame units, and fully enclosed boxes. (Courtesy of Computer Products, Inc.)

inside a larger instrument. They come in a wide range of voltages and currents and include dual and triple units as well as single-output supplies. For example, a popular triple open-frame linear provides +5 volts at 3 amps and ± 15 volts at 0.8 amp and costs \$75 in small quantity. Open-frame supplies are larger than potted modules, and you always screw them to the chassis. Open-frame linears fall in the 10–200 watt range, switchers in the 20–400 watt range. Open-frame supplies at the

low end of the power range may have all components mounted directly on a circuit board, with no metal frame at all. As with the potted supplies, you are expected to provide switches, filters, and fuses for the ac line voltage circuits.

3. Fully enclosed supplies: These supplies have a full metal enclosure, usually perforated for cooling, and usually free of the protruding power transistors, etc., that you find on an open-frame supply. They can be mounted externally, because their full

enclosure keeps fingers out; you can also mount them inside an instrument, if you want. They come with single and multiple outputs, in both linear and switchers. Fully enclosed linears fall in the 15–750 watt range, switchers in the 25–1500 watt range.

4. Wall plug-in power supplies: These are the familiar black plastic boxes that come with small consumer electronic gadgets and plug directly into the wall. They actually come in three varieties, namely (a) step-down ac transformer only, (b) unregulated dc supply, and (c) complete regulated supply; the latter can be either linear or switcher. For example, Ault has a nice series of dual ($\pm 12\text{V}$ or $\pm 15\text{V}$) and triple ($+5\text{V}$ and $\pm 12\text{V}$ or $\pm 15\text{V}$) linear regulated wall-plug-in supplies. These save you the trouble of bringing the ac line power into your instrument, and keep it light and small. Some of us think that these convenient supplies are getting a bit *too* popular, though, as measured by the cluster of wall plug-ins found feeding at the outlets in our house! Some “desktop” models have two cords, one each for the ac input and dc output. Some of the switching units allow a full 95 to 252 volts ac input range, useful for traveling instruments. We’ll have more to say about wall plug-ins in Section 14.03, when we deal with low-power design.

SELF-EXPLANATORY CIRCUITS

6.26 Circuit ideas

Figure 6.65 presents a variety of current ideas, mostly taken from manufacturers’ data sheets.

6.27 Bad circuits

Figure 6.66 presents some circuits that are guaranteed not to work. Figure them out, and you will avoid these pitfalls.

ADDITIONAL EXERCISES

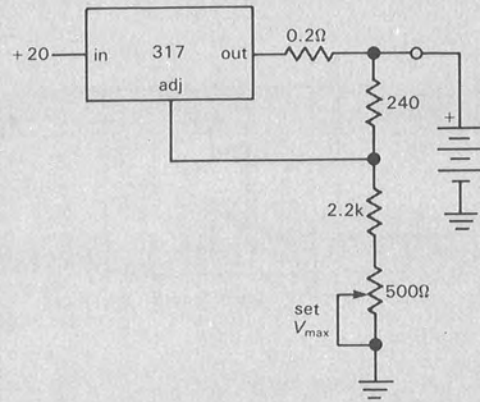
(1) Design a regulated supply to deliver exactly $+10.0$ volts at currents up to 10mA using a 723. You have available a 15 volt (rms) 100mA transformer, diodes by the bucketful, various capacitors, a 723, resistors, and a 1k trimmer pot. Choose your resistors so that they are standard (5%) values and so that the range of adjustment of the trimmer will be sufficient to accommodate the production spread of internal reference voltages (6.80V to 7.50V).

(2) Design $+5$ volt 50mA voltage regulators, assuming $+10$ volt unregulated input, using the following: (a) zener diode plus emitter follower; (b) 7805 3-terminal regulator; (c) 723 regulator; (d) 723 plus outboard *npn* pass transistor; use foldback current limiting with 100mA onset (full-voltage current limit) and 25mA short-circuit current limit; (e) a 317 3-terminal adjustable positive regulator; (f) discrete components, with zener reference and feedback. Be sure to show component values; provide 100mA current limiting for (a), (c), and (f).

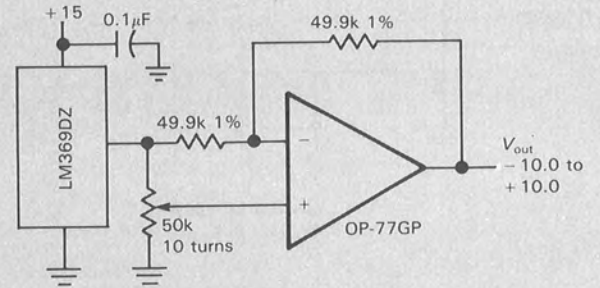
(3) Design a complete $+5$ volt 500mA power supply for use with digital logic. Begin at the beginning (the 115V ac wall socket), specifying such things as transformer voltage and current ratings, capacitor values, etc. To make your job easy, use a 7805 3-terminal regulator. Don’t squander excess capacitance, but make your design conservative by allowing for $\pm 10\%$ variation in all parameters (power-line voltage, transformer and capacitor tolerances, etc.). When you’re finished, calculate worst-case dissipation in the regulator.

Next, modify the circuit for 2 amp load capability by incorporating an outboard pass transistor. Include a 3 amp current limit.

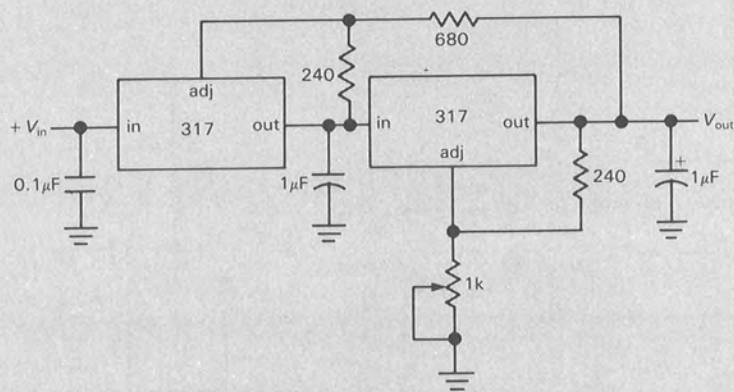
Circuit ideas



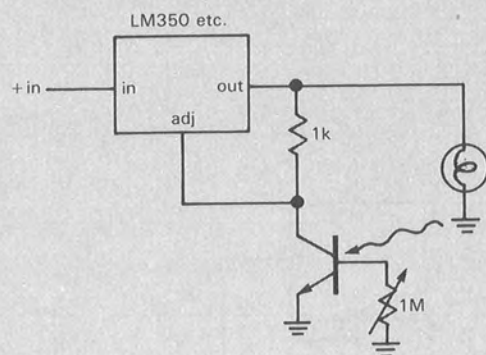
A. 12V battery charger



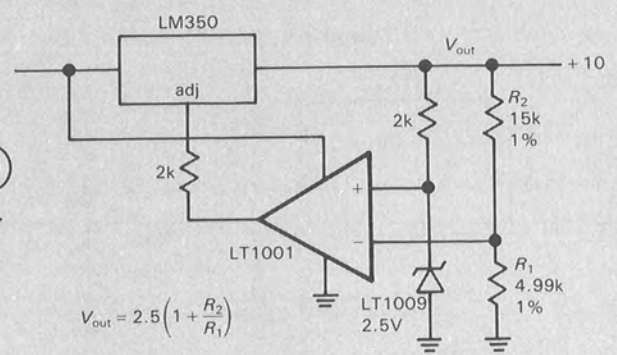
B. adjustable stable bipolar voltage reference



C. tracking preregulator



D. automatic incandescent bulb light regulator

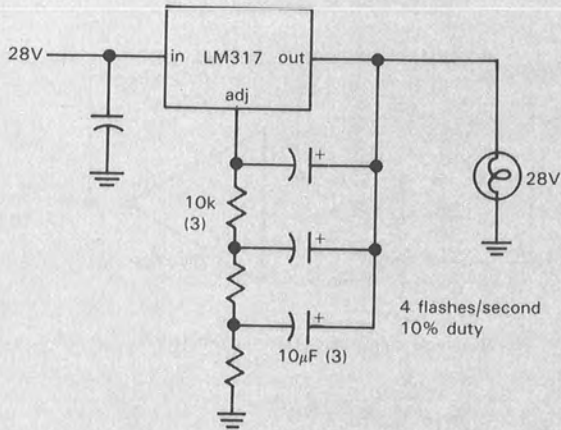


E. precision power voltage source

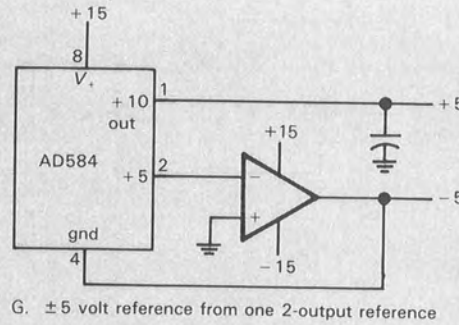
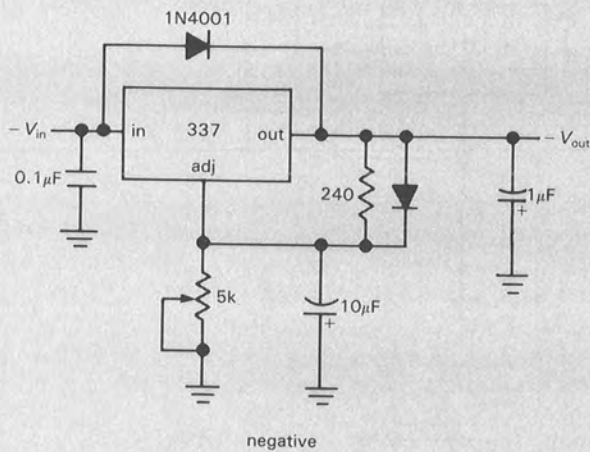
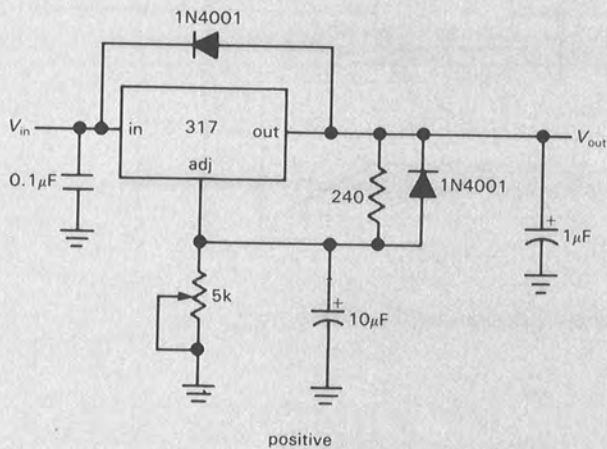
Figure 6.65



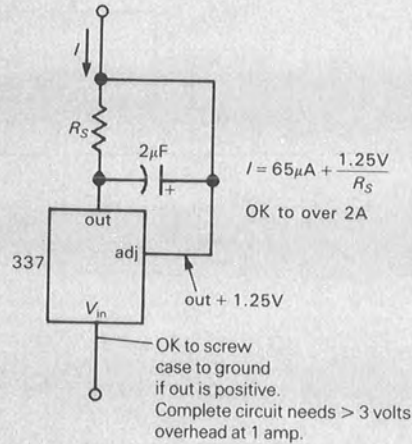
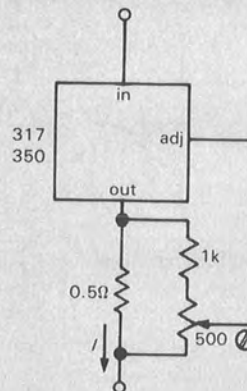
Circuit ideas (cont.)



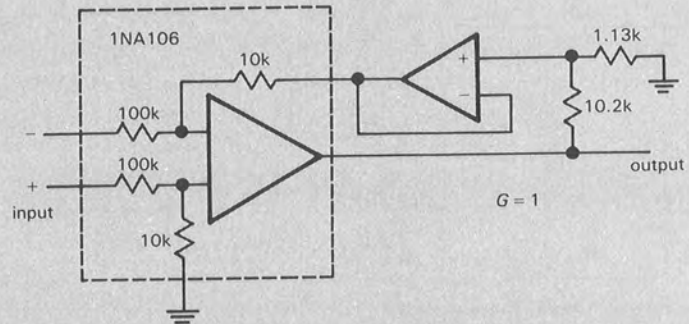
F. lamp flasher (from NSC317 data sheet)



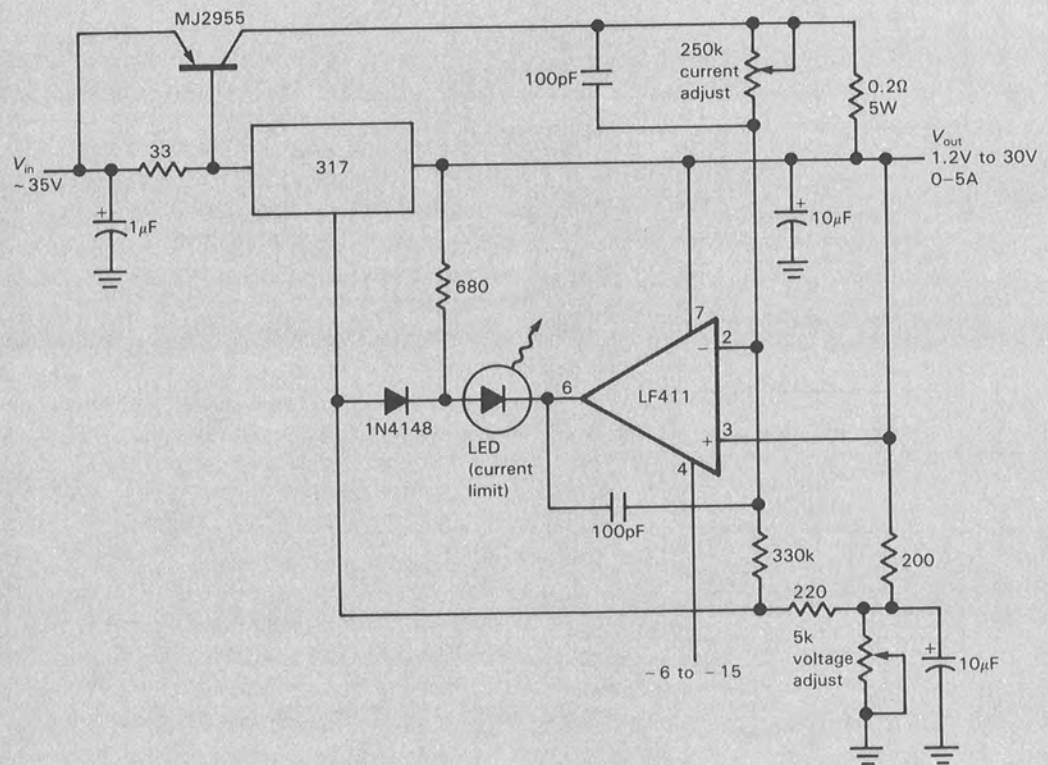
H. 3-terminal regulators with improved ripple rejection (diodes protect against input/output shorts)



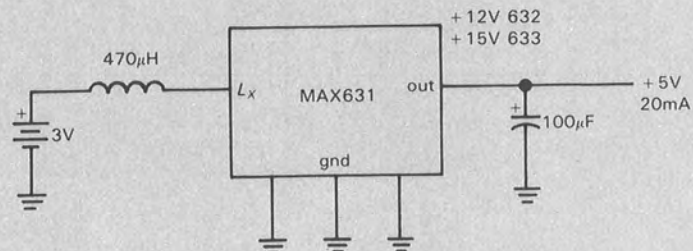
I. power current sources



J. $\pm 100\text{V}$ common-mode range differential follower



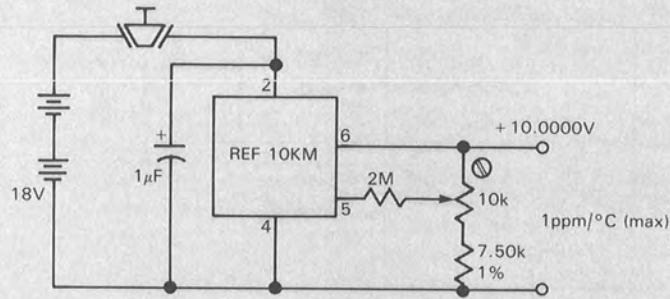
K. constant voltage/constant current supply



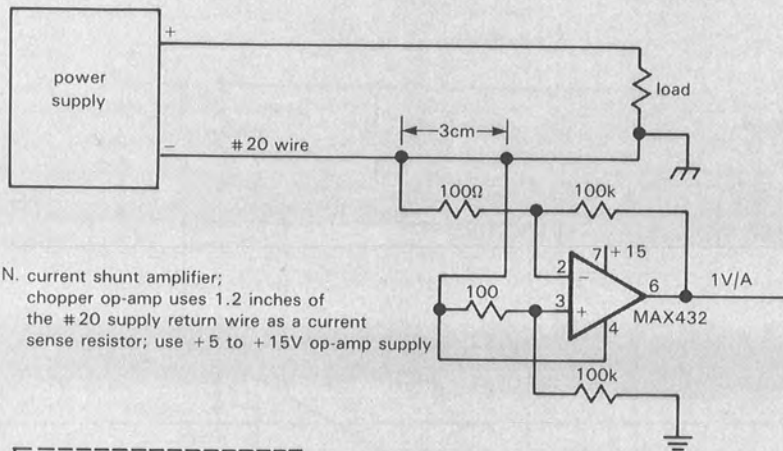
L. "world's simplest" dc-dc converter



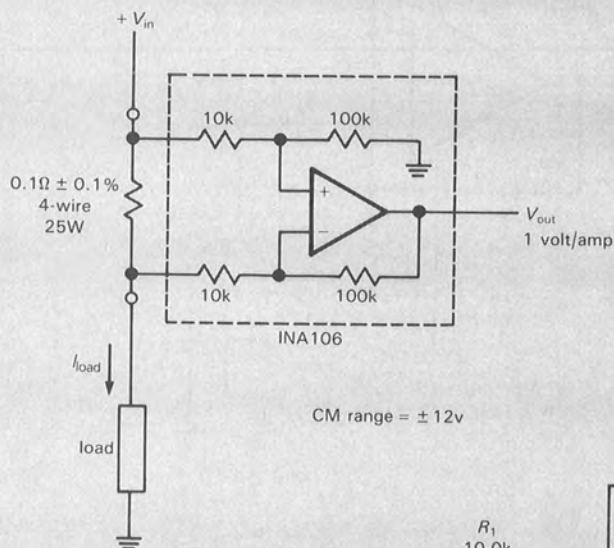
Circuit ideas (cont.)



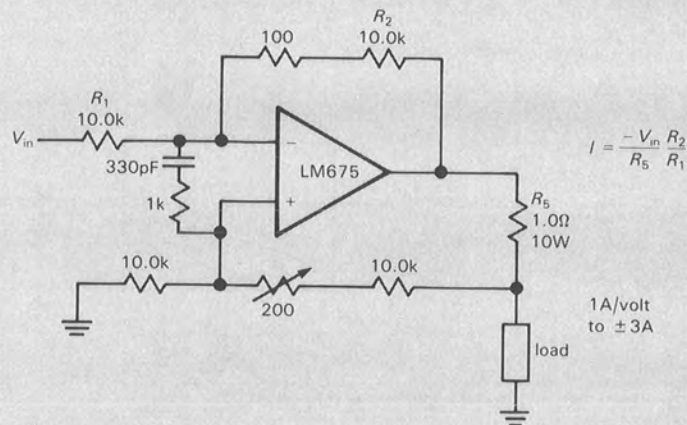
M. portable 10ppm voltage reference



N. current shunt amplifier;
chopper op-amp uses 1.2 inches of
the #20 supply return wire as a current
sense resistor; use +5 to +15V op-amp supply



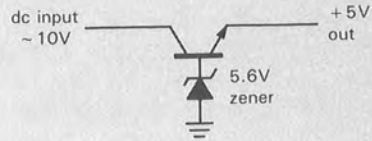
O. high-side current monitor



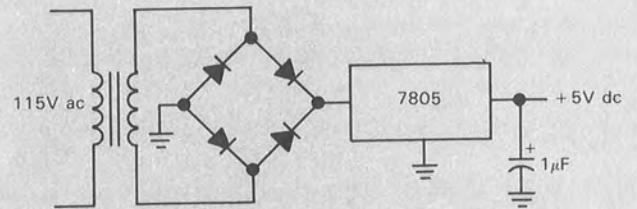
P. high-current bipolarity current source



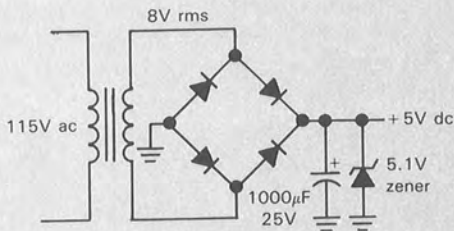
Bad circuits



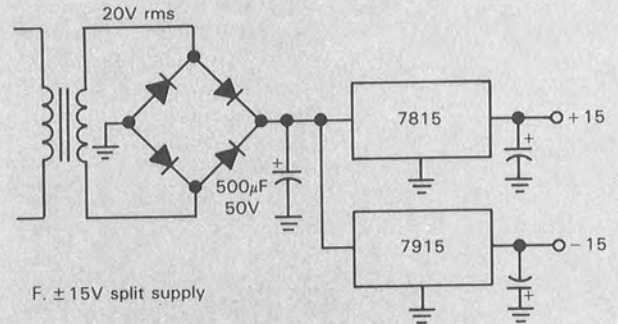
A. simple regulated supply



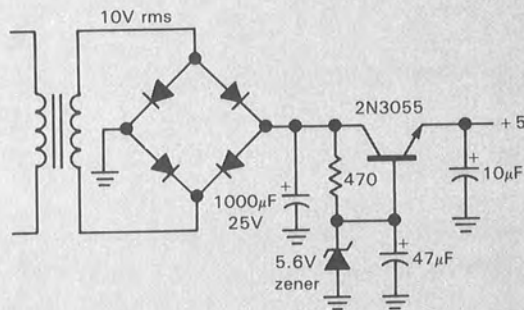
E. +5V supply



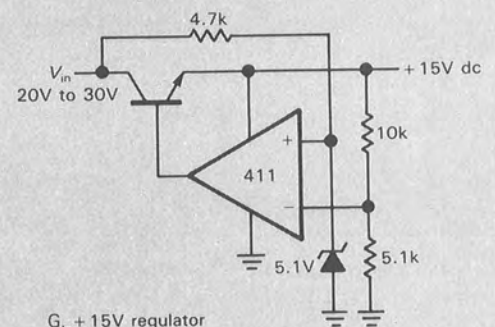
B. +5V supply



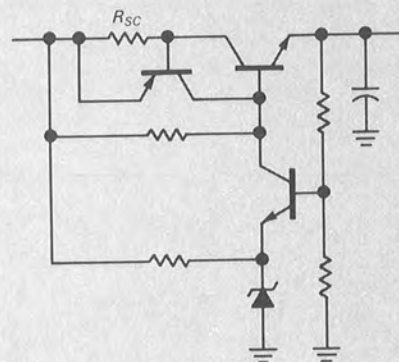
F. ±15V split supply



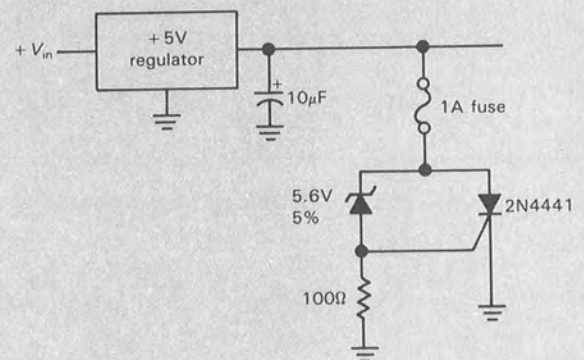
C. +5V supply



G. +15V regulator



D. regulator with upstream current limiter



H. crowbar with SCR protection

Figure 6.66

PRECISION CIRCUITS AND LOW-NOISE TECHNIQUES

CHAPTER

7

In the preceding chapters we have dealt with many aspects of analog circuit design, including the circuit properties of passive devices, transistors, FETs, and op-amps, the subject of feedback, and numerous applications of these devices and circuit methods. In all our discussions, however, we have not yet addressed the question of the best that can be done, for example, in minimizing amplifier errors (non-linearities, drifts, etc.) and in amplifying weak signals with minimum degradation by amplifier “noise.” In many applications these are the most important issues, and they form an important part of the art of electronics. In this chapter, therefore, we will look at methods of precision circuit design and the issue of noise in amplifiers. With the exception of the introduction to noise in Section 7.11, this chapter can be skipped over in a first reading. This material is not essential for an understanding of later chapters.

PRECISION OP-AMP DESIGN TECHNIQUES

In the field of measurement and control there is often a need for circuits of high precision. Control circuits should be accurate, stable with time and temperature, and predictable. The usefulness of measuring instruments likewise depends on their accuracy and stability. In almost all electronic subspecialties we always have the desire to do things more accurately – you might call it the joy of perfection. Even if you don’t always actually *need* the highest precision, you can still delight in the joy of fully understanding what’s going on.

7.01 Precision versus dynamic range

It is easy to get confused between the concepts of *precision* and *dynamic range*, especially since some of the same techniques are used to achieve both. Perhaps the difference can best be clarified by some examples: A 5-digit multimeter has high

precision; voltage measurements are accurate to 0.01% or better. Such a device also has wide dynamic range; it can measure millivolts and volts on the same scale. A precision decade amplifier (one with selectable gains of 1, 10, and 100, say) and a precision voltage reference may have plenty of precision, but not necessarily much dynamic range. An example of a device with wide dynamic range but only moderate accuracy might be a 6-decade logarithmic amplifier (log amp) built with carefully trimmed op-amps but with components of only 5% accuracy; even with accurate components a log amp might have limited accuracy because of lack of log conformity (at the extremes of current) of the transistor junction used for the conversion. Another example of a wide-dynamic-range instrument (greater than 10,000:1 range of input currents) with only moderate accuracy (1%) is the coulomb meter described in Section 9.26. It was originally designed to keep track of the total charge put through an electrochemical cell, a quantity that needs to be known only to approximately 5% but that may be the cumulative result of a current that varies over a wide range. It is a general characteristic of wide-dynamic-range design that input offsets must be carefully trimmed in order to maintain good proportionality for signal levels near zero; this is also necessary in precision design, but, in addition, precise components, stable references, and careful attention to every possible source of error must be used to keep the sum total of all errors within the so-called error budget.

7.02 Error budget

A few words on *error budgets*. There is a tendency for the beginner to fall into the trap of thinking that a few strategically placed precision components will result in a device with precision performance. On rare occasions this will be true. But even

a circuit peppered with 0.01% resistors and expensive op-amps won't perform to expectations if somewhere in the circuit there is an input offset current multiplied by a source resistance that gives a voltage error of 10mV, say. With almost any circuit there will be errors arising all over the place, and it is essential to tally them up, if for no other reason than to locate problem areas where better devices or a circuit change might be needed. Such an error budget results in rational design, in many cases revealing where an inexpensive component will suffice, and eventually permitting a careful estimate of performance.

7.03 Example circuit: precision amplifier with automatic null offset

In order to motivate the discussion of precision circuits, we have designed an extremely precise decade amplifier with automatic offset. This gadget lets you "freeze" the value of the input signal, amplifying any subsequent changes from that level by gains of exactly 10, 100, or 1000. This might come in particularly handy in an experiment in which you wish to measure a small change in some quantity (e.g., light transmission or radiofrequency absorption) as some condition of the experiment is varied. It is ordinarily difficult to get accurate measurements of small changes in a large dc signal, owing to drifts and instabilities in the amplifier. In such a situation a circuit of extreme precision and stability is required. We will describe the design choices and errors of this particular circuit in the framework of precision design in general, thus rendering painless what could otherwise become a tedious exercise. A note at the outset: Digital techniques offer an attractive alternative to the purely analog circuitry used here. Look forward to exciting revelations in chapters to come! Figure 7.1 shows the circuit.

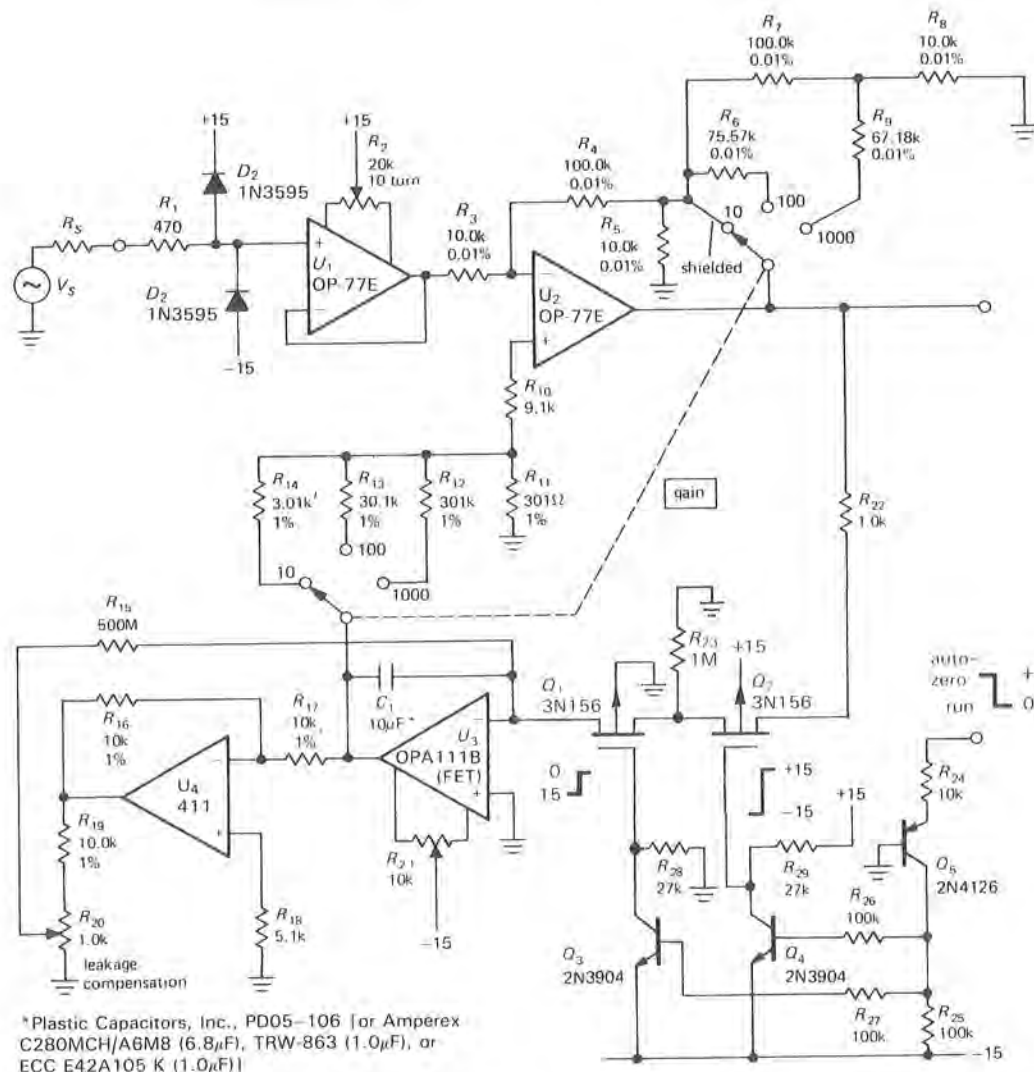


Figure 7.1. Autonulling dc laboratory amplifier.

Circuit description

The basic circuit is a follower (U_1) driving an inverting amplifier of selectable gain (U_2), the latter offsettable by a signal applied to its noninverting input. Q_1 and Q_2 are FETs, used in this application as simple analog switches; Q_3 – Q_5 generate suitable levels, from a logic-level input, to activate the switches. Q_1 through Q_5 and their associated circuitry could all be replaced by a relay, or even a switch,

if desired. For now, just think of it as a simple SPST switch.

When the logic input is HIGH (“auto-zero”), the switch is closed, and U_3 charges the analog “memory” capacitor (C_1) as necessary to maintain zero output. No attempt is made to follow rapidly changing signals, since in the sort of application for which this was designed the signals are essentially dc, and some averaging is a desirable feature. When the switch is opened, the voltage on the capacitor

remains stable, resulting in an output signal proportional to the wanderings of the input thereafter.

There are a few additional features that should be described before going on to explain in detail the principles of precision design as applied here: (a) U_4 participates in a first-order leakage-current compensation scheme, whereby the tendency of C_1 to discharge slowly through its own leakage (100,000M, minimum, corresponding to a time constant of 2 weeks!) is compensated by a small charging current through R_{15} proportional to the voltage across C_1 . (b) Instead of a single FET switch, two are used in series in a “guarded leakage-cancellation” arrangement. The small leakage current through Q_2 , when switched OFF, flows to ground through R_{23} , keeping all terminals of Q_1 within millivolts of ground. Without any appreciable voltage drops, Q_1 hasn’t any appreciable leakage! (See Section 4.15 and Fig. 4.50 for similar circuit tricks.) (c) The offsetting voltage generated at the output of U_3 is attenuated by R_{11} – R_{14} , according to the gain setting. This is done to avoid problems with dynamic range and accuracy in U_3 , since drifts or errors in the offset holding circuitry are not amplified by U_2 (more on this later).

7.04 A precision-design error budget

For each category of circuit error and design strategy we will devote a few paragraphs to a general discussion, followed by illustrations from the preceding circuit. Circuit errors can be divided into the categories of (a) errors in the external network components, (b) op-amp (or amplifier) errors associated with the input circuitry, and (c) op-amp errors associated with the output circuitry. Examples of the three are resistor tolerances, input offset voltage, and errors due to finite slew rate, respectively.

Let’s start by setting out our error budget. It is based on a desire to keep input errors down to the $10\mu\text{V}$ level, output drift (from capacitor “droop”) below 1mV in 10 minutes, and gain accuracy in the neighborhood of 0.01%. As with any budget, the individual items are arrived at by a process of trade-offs, based on what can be done with available technology. In a sense the budget represents the end result of the design, rather than the starting point. However, it will aid our discussion to have it now.

Error budget (worst-case values)

1. Buffer amplifier (U_1)

Voltage errors referred to input:

Temperature	$1.2\mu\text{V}/4^\circ\text{C}$
Time	$1.0\mu\text{V}/\text{month}$
Power supply	$0.3\mu\text{V}/100\text{mV change}$
Bias current $\times R_S$	$2.0\mu\text{V}/1\text{k of } R_S$
Load-current heating	$0.3\mu\text{V @ full scale (10V)}$

2. Gain amplifier (U_2)

Voltage errors referred to input:

Temperature	$1.2\mu\text{V}/4^\circ\text{C}$
Time	$1.0\mu\text{V}/\text{month}$
Power supply	$0.3\mu\text{V}/100\text{mV change}$
Bias offset current drift	$1.6\mu\text{V}/4^\circ\text{C}/1\text{k}$
Load-current heating	$0.3\mu\text{V @ full scale}$ ($R_L \geq 10\text{k}$)

3. Hold amplifier (U_3)

Voltage errors referred to output:

U_3 offset tempco	$60\mu\text{V}/4^\circ\text{C}$
Power supply	$10\mu\text{V}/100\text{mV change}$
Capacitor droop	$100\mu\text{V}/\text{min}$
(see current error budget)	
Charge transfer	$10\mu\text{V}$

Current errors applied to C_1 (needed for preceding voltage error budget):

Capacitor leakage	
Maximum (uncompensated)	(100pA)
Typical (compensated)	10pA
U_3 input current	0.2pA
U_3 & U_4 offset voltage	
across R_{15}	1.0pA
FET switch OFF leakage	0.5pA
Printed-circuit-board leakage	5.0pA

The various items in the budget will make sense as we discuss the choices faced in

this particular design. We will organize by the categories of circuit errors listed earlier: network components, amplifier input errors, and amplifier output errors.

7.05 Component errors

The degrees of precision of reference voltages, current sources, amplifier gains, etc., all depend on the accuracy and stability of the resistors used in the external networks. Even where precision is not involved directly, component accuracy can have significant effects, e.g., in the common-mode rejection of a differential amplifier made from an op-amp (see Section 4.09), where the ratios of two pairs of resistors must be accurately matched. The accuracy and linearity of integrators and ramp generators depend on the properties of the capacitors used, as do the performances of filters, tuned circuits, etc. As you will see shortly, there are places where component accuracy is crucial, and there are other places where the particular component value hardly matters at all.

Components are generally specified with an initial accuracy, as well as the changes in value with time (stability) and temperature. In addition, there are specifications of voltage coefficient (nonlinearity) and bizarre effects such as "memory" and dielectric absorption (for capacitors). Complete specifications will also include the effects of temperature cycling and soldering, shock and vibration, short-term overloads, and moisture, with well-defined conditions of measurement. In general, components of greater initial accuracy will have their other specifications correspondingly better, in order to provide an overall stability comparable with the initial accuracy. However, the overall error due to all other effects combined can exceed the initial accuracy specification. Beware!

As an example, RN55C 1% tolerance metal-film resistors have the following

specifications: temperature coefficient (tempco), 50ppm/°C over the range -55°C to $+175^{\circ}\text{C}$; soldering, temperature, and load cycling, 0.25%; shock and vibration, 0.1%; moisture, 0.5%. By way of comparison, ordinary 5% carbon-composition resistors (Allen-Bradley type CB) have these specifications: tempco, 3.3% over the range 25°C to 85°C ; soldering and load cycling, +4%, -6%; shock and vibration, $\pm 2\%$; moisture, +6%. From these specs it should be obvious why you can't just select (using an accurate digital ohmmeter) carbon resistors that happen to be within 1% of their marked value for use in a precise circuit, but are obliged to use 1% resistors (or better) designed for long-term stability as well as initial accuracy. For the utmost in precision it is necessary to use an ultra-precise metal-film resistor, such as Mepco 5023Z (5ppm/°C and 0.025%), or wire-wound resistors, available with tolerances of 0.01%. See Appendix D for more information on precision resistors.

Nulling amplifier: component errors

In the preceding circuit (Fig 7.1), 0.01% resistors are used in the gain-setting network, R_3 – R_9 , giving highly predictable gain. As you will see shortly, the value of R_3 is a compromise, with small values reducing offset current error in U_2 but increasing heating and thermal offsets in U_1 . Given the value of R_3 , the feedback network is forced to take on its complicated form to keep the resistor values below 301k, the maximum value generally available in 1% precision resistors. This trick is discussed in Section 4.19. Note that 1% resistors are used in the offset attenuator network, R_{11} – R_{14} ; here accuracy is irrelevant, and metal-film resistors are used only for their good stability.

The largest error term in this circuit, as the error budget shows, is capacitor leakage in the holding capacitor, C_1 . Capacitors intended for low-leakage applications

give a leakage specification, sometimes as a leakage resistance, sometimes as a time constant (megohm-microfarads). In this circuit C_1 must have a value of at least a few microfarads in order to keep the charging rate from other current error terms small (see budget). In that range of capacitance, polystyrene, polycarbonate, and polysulfone capacitors have the lowest leakage. The unit chosen has a leakage specified as 1,000,000 megohm-microfarads maximum, i.e., a parallel leakage resistance of at least 100,000M. Even so, that's equivalent to a leakage current of 100pA at full output (10V), corresponding to a droop rate of nearly 1mV/min at the output, the largest error term by far. For that reason we have added the leakage-cancellation scheme described earlier. It is fair to assume that the effective leakage can be reduced to 0.1 of the capacitor's worst-case leakage specification (in practice, we can probably do much better). No great stability is required in the cancellation circuit, given the modest demands made of it. As you will see later when we discuss voltage offsets, R_{15} is kept intentionally large so that input voltage offsets in U_3 aren't converted to a significant current error.

While on the subject of errors produced by components external to the amplifiers themselves, it should be pointed out that leakage in FET switches is normally in the range of 1nA, a value completely unacceptable in this circuit. The trick of using a pair of series-connected FETs, with Q_2 's leakage resulting in only 1mV across Q_1 (with negligible leakage into U_3 's summing junction), is elegant and powerful; it is sometimes used in integrator circuits, as discussed in Section 4.19. We have also used it in a novel peak-detector circuit in Section 4.15. As you will see shortly, U_3 is chosen carefully to keep currents through C_1 down in the picoampere range. The philosophy is

the same everywhere: Choose circuit configurations and component types as necessary to meet the error budget. At times this involves hard work and circuit trickery, but at other times it falls easily within standard practice.

One subtle source of error in any circuit using FET switches is *charge transfer* from the controlling gate to the signal-carrying channel. The full-swing transitions at the gate couple capacitively to the drain and source. As we remarked in Chapter 3, the total charge transferred is independent of the transition time and depends only on the gate swing and gate-channel capacitance: $\Delta Q = C_{GC}\Delta V_G$. In this circuit, charge transfer results in a simple voltage error of the auto-zero, because the charge is converted to a voltage in the holding capacitor C_1 . It's easy to estimate the error. The 3N156 specifies a C_{rss} (drain-gate capacitance) of 1.3pF maximum, and a C_{iss} (gate-channel capacitance, mostly to the source) of 5pF maximum. The 15 volt gate swing therefore produces a maximum charge transfer of 75pC, corresponding to a voltage step of $\Delta V_C = \Delta Q/C_1 = 7.5\mu\text{V}$ across the 10 μF capacitor C_1 . This is within our error budget; in fact, we may have overestimated the effect, since we included the capacitance to source as well as drain, whereas during a portion of the gate step the channel is cut off, decoupling the source from the drain.

7.06 Amplifier input errors

The deviations of op-amp input characteristics from the ideal that we discussed in Chapter 4 (finite values of input impedance and input current, voltage offset, common-mode rejection ratio, and power-supply rejection ratio, and their drifts with time and temperature) generally constitute serious obstacles to precision circuit design and force trade-offs in circuit configuration, component selection, and the choice of a particular

op-amp. The point is best made with examples, as we will do shortly. Note that these errors, or their analogs, exist for amplifiers of discrete design as well.

Input impedance

Let's discuss briefly the error terms just listed. The effect of finite input impedance is to form voltage dividers in combination with the source impedance driving the amplifier, reducing the gain from the calculated value. Most often this isn't a problem, because the input impedance is bootstrapped by feedback, raising its value enormously. As an example, the OP-77E precision op-amp (with transistor, not FET, input stage) has a typical "differential-mode input impedance," of 45M. In a circuit with plenty of loop gain, feedback raises the input impedance to the "common-mode input impedance" 200,000M. In any case, some FET-input op-amps have astronomical values of R_{in} , if there's still a problem.

Input bias current

More serious is the input bias current. Here we're talking about currents measured in nanoamps, and this already produces voltage errors of microvolts for source impedances as small as 1k. Again, FET op-amps come to the rescue, but with generally increased voltage offsets as part of the bargain. Bipolar superbeta op-amps such as the LT1012, 312, and LM11 can also have surprisingly low input currents. As an example, compare the OP-77 precision bipolar op-amp with the LT1012 (bipolar, optimized for low bias current), the OPA111 (JFET, precision and low bias), the AD549 (ultra-low-bias JFET), and the ICH8500 (MOSFET, lowest-bias op-amp); these are the best you can get at the time of

writing, and we've chosen the best grade of each one:

	Bias current @ 25°C I_B max	Offset voltage @ 25°C V_{os} max	Tempco of V_{os} ΔV_{os} max
OP-77E (bipolar)	2000pA	25μV	0.3μV/°C
LT1012C (superbeta)	150pA	50μV	1.5μV/°C
OPA111B (JFET)	1pA	250μV	1μV/°C
AD549L (JFET)	0.06pA	500μV	10μV/°C
ICH8500A (MOSFET)	0.01pA	50,000μV	2000μV/°C

Well-designed FET amplifiers have extremely low bias current, but with much larger offset voltage, as compared with the precision OP-77. Since the offset voltage can always be trimmed, what matters more is the drift with temperature. In this case the FET amplifiers are 3 to 6000 times worse. The op-amp with the lowest input current uses MOSFETs for the input stage. MOSFET op-amps are becoming popular because of the availability of inexpensive units like the 3440, 3160, the TLC270 series, and the ICL7610 series, as well as the ultra-low-bias-current devices like the 8500A listed earlier. However, unlike JFETs or bipolar transistors, MOSFETs can have very large drifts of offset voltage with time, an effect that will be discussed shortly. So the improvement in current errors you buy with a FET op-amp can be wiped out by the larger voltage error terms. With any circuit in which bias current can contribute significant error, it is always wise to ensure that both op-amp input terminals see the same dc source resistance, as we discussed in Section 4.12; then the op-amp's *offset current* becomes the relevant specification. A note on bias-current compensation: A number of precision op-amps use a "bias compensation" scheme to cancel (approximately) the input current, in order to make that error

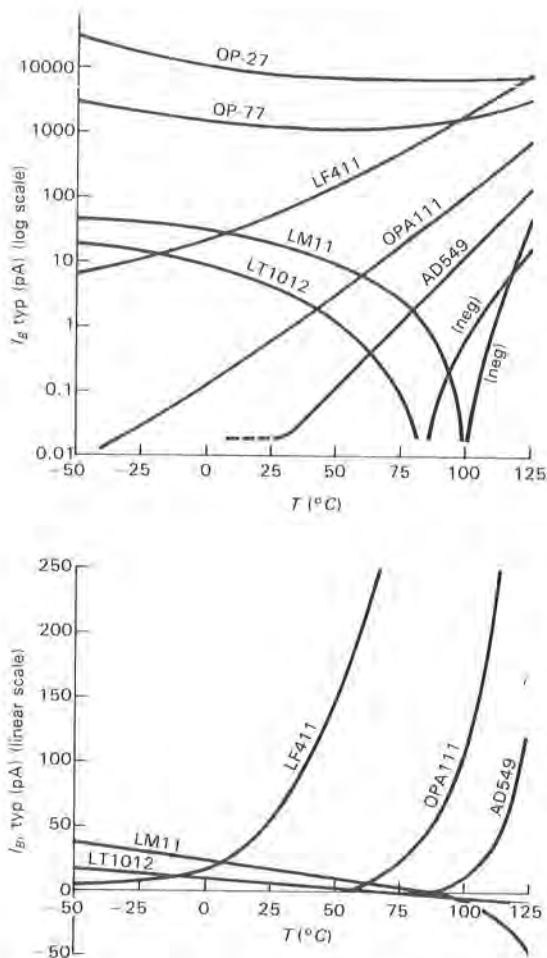


Figure 7.2. Op-amp input current versus temperature.
A. Logarithmic scale
B. Linear scale

term smaller; look back at Additional Exercise 8 at the end of Chapter 2 to see how it's done. With op-amps of this type you generally don't gain anything by matching the dc resistances seen by the two inputs, since for a bias-compensated op-amp the residual bias current and the offset current are comparable.

One additional point to keep in mind when using FET-input op-amps is that the input 'bias' current is actually gate leakage current, and it rises dramatically with increasing temperature (it roughly doubles

for every 10°C increase in chip temperature; see Fig. 3.30). Since FET op-amps often run warm (the popular 356 dissipates 150mW quiescent power), the actual input current may be considerably higher than the 25°C figures you see on the data sheet. The input current of a bipolar-transistor-input op-amp, by comparison, is actual base current, and it *drops* with rising temperature (Fig. 7.2). So a FET-input op-amp with impressive input-current specs on paper may not give such an improvement over a good superbeta bipolar unit. As an example, the OPA111 with its 1pA input current (at 25°C) will have an input current of about 10pA at 65°C chip temperature, which is higher than the input current of the superbeta LT1012 at the same temperature. The popular 355 series of FET op-amps has an input current that is comparable to that of the LT1012 or LM11 at 25°C and is many times higher at elevated temperatures. Finally, when comparing op-amp input currents, watch out for some FET types whose I_B depends on the input voltage. The spec sheet usually lists I_B only at 0 volts (mid-supply), but a good data sheet will show curves as well. See Figure 7.3 for some typical I_B - V_{in} behavior. Note the excellent performance of the OPA111, due in part to its cascode input stage.

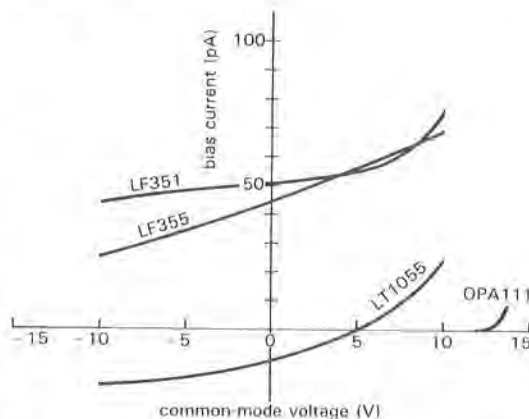


Figure 7.3. FET op-amp input current versus common-mode voltage.

Voltage offset

Voltage offsets at the amplifier input are obvious sources of error. Op-amps differ widely in this parameter, ranging from “precision” op-amps offering worst-case V_{os} values generally in the tens of microvolts to ordinary jellybean op-amps like the LF411 with V_{os} values of 2 to 5mV. At the time of writing, the champion in the (non-chopper, see below) world of low offsets is the MAX400M ($V_{os} = 10\mu\text{V}$, max). We expect to see further incremental improvements in this area.

Although most good single op-amps (but not duals or quads) have offset-adjustment terminals, it is still wise to choose an amplifier with inherently low initial offset V_{os} max, for several reasons. First, op-amps designed for low initial offset tend to have correspondingly low offset drift with temperature and time. Second, a sufficiently precise op-amp eliminates the need for external trimming components (a trimmer takes up space, needs to be adjusted initially, and may change with time). Third, offset voltage drift and common-mode rejection are degraded by the unbalance caused by an offset-adjustment trimmer.

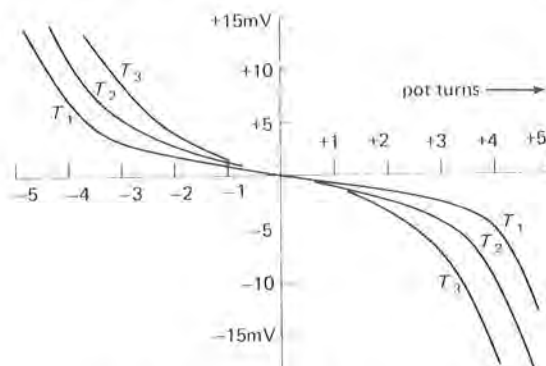


Figure 7.4. Typical op-amp offset versus offset-adjustment potentiometer rotation for several temperatures.

Figure 7.4 illustrates how a trimmed offset has larger drifts with temperature. We

have also shown how the offset adjustment is spread over the trimmer pot rotation, with best resolution near the center, especially for large values of trimmer resistance. Finally, you'll generally find that the recommended external trimming network provides far too much range, making it nearly impossible to trim V_{os} down to a few microvolts; even if you succeed, the adjustment is so critical it won't stay trimmed for long. Another way to think about it is to realize that the manufacturer of a precision op-amp has *already* trimmed the offset voltage, in a custom test jig using “laser-zapping” techniques; you may be unable to do any better yourself. Our advice is (a) to use precision op-amps for precision circuits, and (b) if you must trim them further, arrange a narrow-range trim circuit like the ones shown in Figure 7.5, which have a full-scale range of $\pm 50\mu\text{V}$, linear in trimmer rotation.

Because voltage offsets can be trimmed to zero, what ultimately matters is the drift of offset voltage with time, temperature, and power-supply voltage. Designers of precision op-amps work hard to minimize these errors. You get the best performance from bipolar (as opposed to FET) op-amps in this regard, but input current effects may then dominate the error budget. The best op-amps keep drifts below $1\mu\text{V}/^\circ\text{C}$; at the time of writing, the AD707 claims the smallest drift (for a non-chopper op-amp) — $\Delta V_{os} = 0.1\mu\text{V}/^\circ\text{C}$, max.

Another factor to keep in mind is the drift caused by self-heating of the op-amp when it drives a low-impedance load. It is often necessary to keep the load impedance above 10k to prevent large errors from this effect. As usual, that may compromise the next stage's error budget from the effects of bias current! You will see just such a problem in this design example. For applications where drifts of a few microvolts are important, the related effects of thermal gradients (from nearby heat-producing components) and thermal emf's

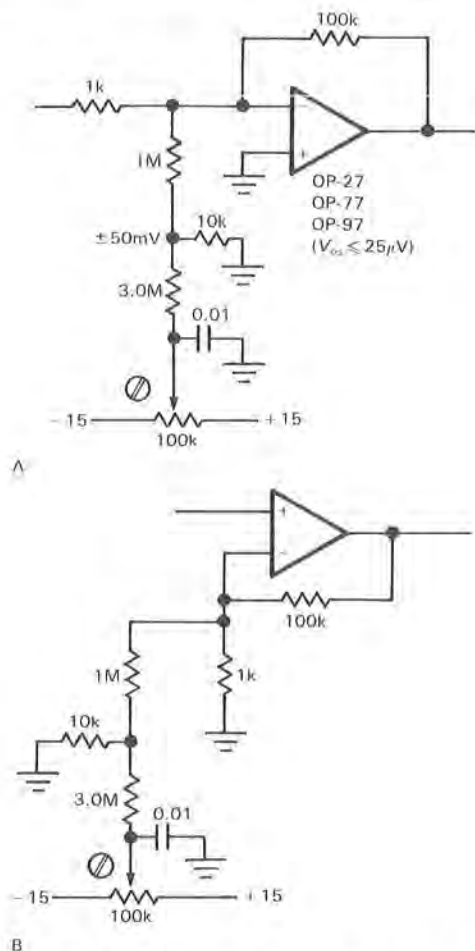


Figure 7.5. External trimming networks for precision op-amps.
A. Inverting.
B. Noninverting.

(from junctions of dissimilar metals) become important. This will come up again when we discuss the ultraprecise *chopper-stabilized* amplifier in Section 7.08.

Table 7.1 compares the important specifications for seven of our favorite precision op-amps. Spend some time with it – it will give you a good feeling for the trade-offs you face in high-performance design with op-amps. Note particularly the trade-offs of offset voltage (and drift) versus input current for the best bipolar and JFET op-amps. You also get the lowest noise voltage from bipolar op-amps, dropping with

increasing bias current; we'll see why that happens later in the chapter, when we discuss noise. The awards for low-noise *current*, however, always go to the FET op-amps, again for reasons that will become clear later. In general, choose FET op-amps for low input current and current noise; choose bipolar op-amps for low input voltage offset, drift, and voltage noise.

Among FET-input op-amps, those using JFETs dominate the scene, particularly where precision is needed. MOSFETs, in particular, are subject to a unique debilitating effect that neither FETs nor bipolar transistors have. It turns out that sodium-ion impurities in the gate insulating layer migrate slowly under the influence of the gate's $V_{GS(ON)}$ electric field, resulting in a drift of the offset voltage under closed-loop conditions of as much as 0.5mV over a period of years. The effect is increased for elevated temperatures and for a large applied differential input signal. For example, the data sheet for the CA3420 MOSFET-input op-amp shows a typical 5mV change of V_{os} over 3000 hours of operation at 125°C with 2 volts across the input. This sodium-ion disease can be cured by introducing phosphorus into the gate region. Texas Instruments, for example, uses a phosphorus-doped polysilicon gate in its "LinCMOS" series of op-amps (TLC270 series) and comparators (TLC339 and TLC370 series). These popular inexpensive parts come in a variety of packages and speed/power selections and maintain respectable offset voltages with time (50μV eventual offset drift per volt of differential input).

There is an important exception to the generalization that FET op-amps, particularly MOSFET types, suffer from larger initial offsets and much larger drifts of V_{os} with temperature and time than do bipolar-transistor op-amps. That exception is the so-called auto-zero (or "chopper-stabilized") amplifier, which uses MOSFET analog switches and amplifiers

Parameter	Symbol	bipolar OP-77E		low-noise OP-27E		low-bias LT1012C		micropower OP-90E		fast JFET LT1055A		low-bias JFET OPA111B		chopper MAX430C		units
		typ	min/max	typ	min/max	typ	min/max	typ	min/max	typ	min/max	typ	min/max	typ	min/max	
$(V_s = \pm 15V; T_A = 25^\circ C)$																
Offset voltage	V_{os}	10	25	10	25	10	50	50	150	50	150	50	250	1	5	μV
Offset voltage drift	ΔV_{os}	0.2	-	0.2	1	0.3	-	0.3	2	-	-	-	-	0.1	-	$\mu V/month$
Bias current	I_b	1.2	2	10	40	0.03	0.15	4	15	0.01	0.05	0.0005	0.001	0.02	0.2	nA
Offset current	I_{os}	0.1	1.5	7	35	0.02	0.15	0.4	3	0.002	0.01	0.0003	0.0008	0.01	0.1	nA
Input resistance – diff ¹	R_{in}	45	25	6	1.3	-	-	30	-	10^6	-	10^7	-	10^6	-	M Ω
Input resistance – common mode	R_{inCM}	200	-	3	-	-	-	20	-	10^3	-	10^5	-	10^3	-	G Ω
Input noise voltage (0.1-10Hz)	e_n pp	0.4	0.6	0.1	0.2	0.5	-	3	-	1.8	-	1.2	2.5	1.1	-	μV , pp
Input noise voltage density (10Hz)	e_n	10	18	3.5	5.5	17	30	60	-	28	50	30	60	-	-	nV/ $\sqrt{Hz}$
Input noise voltage density (1kHz)	e_n	10	11	3	4	14	22	60	-	14	20	7	12	-	-	nV/ $\sqrt{Hz}$
Input noise current density (10Hz)	i_n	0.3	0.8	1.7	4	0.02	-	1.5	-	0.002	0.004	0.0004	-	0.01	-	pA/ $\sqrt{Hz}$
Input noise current density (1kHz)	i_n	0.1	0.2	0.4	0.6	0.006	-	0.7	-	0.002	0.004	0.0004	-	-	-	pA/ $\sqrt{Hz}$
Large signal voltage gain	A_{vo}	12	5	1.8	1	0.2	2	1.2	0.7	0.4	0.015	2	1	30	1	V/ μV
Common-mode rejection ratio	CMRR	140	120	126	114	132	110	130	100	100	86	110	100	140	120	dB
Power supply rejection ratio	PSRR	120	110	120	100	132	110	120	105	106	90	110	100	140	120	dB
Slew rate	SR	0.3	0.1	2.8	1.7	0.2	0.1	0.01	0.005	13	10	2	1	0.5	-	V/ μs
Gain-bandwidth product	GBW	0.6	0.2	8	5	1	-	0.02	-	5	-	2	-	0.5	-	MHz
Supply current	I_s	1.7	2	3	4.7	0.4	0.6	0.014	0.02	2.8	4	2.5	3.5	1.3	2	mA
(Over temperature range)																
Offset voltage	T	-25 to +85	-25 to +85	-25 to +85	-25 to +85	0 to +70	-25 to +85	-25 to +85	-25 to +85	0 to +70	-25 to +85	-25 to +85	-25 to +85	0 to +70	-	$^\circ C$
Offset voltage tempco	V_{os}	10	45	20	50	20	120	70	270	100	330	100	500	-	-	μV
Bias current	TCV_{os}	0.1	0.3	0.2	0.6	0.2	1.5	0.3	2	1.2	4	0.5	1	0.02	0.05	$\mu V/^\circ C$
Offset current	I_b	2.4	4	14	60	0.035	0.23	4	15	0.03	0.15	0.03	0.13	0.05	-	nA
	I_{os}	0.1	2.2	10	50	0.02	0.23	0.8	3	0.01	0.05	0.02	0.1	0.04	-	nA

to sense, and correct, the residual offset error of an ordinary op-amp (which itself is often built with MOSFETs, on the same chip). Chopper-stabilized op-amps deliver lower voltage offsets and drifts than even the best precision bipolar op-amps – $5\mu\text{V}$ (max), $0.05\mu\text{V}/^\circ\text{C}$ (max) – but not without cost. They have some unpleasant characteristics that make them unsuited for many applications. We will discuss them in detail in Section 7.08.

Common-mode rejection

Insufficient common-mode rejection ratio (CMRR) degrades circuit precision by effectively introducing a voltage offset as a function of dc level at the input. This effect is usually negligible, since it is equivalent to a small gain change, and in any case it can be overcome by choice of configuration: An inverting amplifier is insensitive to op-amp CMRR, in contrast with a noninverting amplifier. However, in “instrumentation amplifier” applications you are looking at a small differential signal riding on a large dc offset, and a high CMRR is essential. In such cases you have to be careful about circuit configurations and, in addition, choose an op-amp with a high CMRR specification. Once again, a superior op-amp like the OP-77 can solve your problems, with a CMRR (min) of 120dB, compared with the 411’s meager specification of 70dB. We will discuss high-gain differential and instrumentation amplifiers shortly.

Power-supply rejection

Changes in power-supply voltage cause small op-amp errors. As with most op-amp specifications, the power-supply rejection ratio (PSRR) is referred to a signal at the *input*. For example, the OP-77 has a specified PSRR of 110dB at dc, meaning that a 0.3 volt change in one of the power-supply voltages causes a change

at the output equivalent to a change in differential input signal of $1\mu\text{V}$.

The PSRR drops drastically with increasing frequency, and a graph documenting this scurrilous behavior is often given on the data sheet. For example, the PSRR of our favorite OP-77 begins dropping at 0.3Hz and is down to 83dB at 60Hz and 42dB at 10kHz. This actually doesn’t present much of a problem, since power-supply noise is also decreasing at higher frequencies if you have used good bypassing. However, 120Hz ripple could present a problem if an unregulated supply is used.

It is worth noting that the PSRR will not, in general, be the same for the positive and negative supplies. Thus, the use of dual-tracking regulators (Section 6.19) doesn’t necessarily bring any benefits.

Nulling amplifier: input errors

The amplifier circuit in Figure 7.1 begins with a follower, to keep a high input impedance. It is tempting to consider a FET type, but the poor V_{os} specification more than offsets the advantage of low input current, except with sources of very high impedance. The OP-77’s 2nA bias current gives an error of $2\mu\text{V}/1\text{k}$ source impedance; a JFET LT1055A, although giving negligible current error, would give voltage offset drifts of $16\mu\text{V}/4^\circ\text{C}$ (4°C is considered a typical laboratory ambient temperature range). The input follower is provided with offset trimming, since the initial $25\mu\text{V}$ spec is too large. As mentioned earlier, feedback bootstraps the input impedance to 200,000M and eliminates any errors from finite source impedance, up to 20M (for gain error less than 0.01%). D_1 and D_2 are included for input overvoltage protection and are low-leakage types (less than 1nA).

U_1 drives an inverting amplifier (U_2), with R_3 being a compromise between heat-produced thermal offsets in U_1 and bias-current offset errors in U_2 . The value

chosen keeps heating down to 5.6mW (at 7.5V output, the worst case), which works out to a temperature rise of 0.8°C (the op-amp has a thermal resistance of about 0.14°C/mW , see Section 6.04), with a consequent voltage offset of $0.3\mu\text{V}$. The resultant 10k source impedance seen by U_2 results in an error due to bias-current offset, but since U_2 is inside a feedback loop with U_3 trimming the overall offset to zero, all that matters is the drift in the current error term. The OP-77 has a specification for bias offset drift with temperature (not often specified by manufacturers, incidentally), from which the error result of $1.6\mu\text{V}/4^{\circ}\text{C}$ in the error budget is calculated. Reducing the value of R_3 would improve this term, at the expense of the heating term in U_1 .

As explained in the overall circuit description earlier, the value of R_3 forces the bizarre feedback T network in order to keep the feedback resistor values in the range where precision wire-wound resistors can be manufactured. Using the ordinary inverting amplifier configuration, for example, you would need resistors of 100.0k, 1.0M, and 10.0M for gains of 10, 100, and 1000, respectively.

The dc input impedance of U_2 comes closer to presenting a problem. At a gain of 1000 its differential input impedance of 25M is bootstrapped by a factor of $A_{\text{vol}}/1000$ to 125,000M. Fortunately this exceeds the 9.4k impedance of the gain-setting network by a factor of more than a million, contributing much less than 0.01% error. This is one of the toughest examples we could think of, and even so the op-amp input impedance presents no problem, thus demonstrating that, in general, you can ignore the effects of op-amp input impedances.

Drifts in offset voltage in both U_1 and U_2 over time, temperature, and power-supply variations affect the final error equally and are tabulated in the budget. It is worth pointing out that they are all

automatically cancelled at each "zeroing" cycle, and only short-term drifts matter anyway. These errors are all in the microvolt range, thanks to a good op-amp. U_3 has larger drifts, but it must be a FET type to keep capacitor current small, as already explained. Since U_3 's output is attenuated according to the gain selected, its error, *referred to the input*, is reduced at high gain. This is an important point, since high gains are used with small input signal levels where high accuracy is needed. U_3 's errors are always the same *at the output*, and they are therefore specified as output errors in the error budget.

Note the general philosophy of design that emerges from this example: You work at the problem areas, choosing configurations and components as necessary to reduce errors to acceptable values. Trade-offs and compromises are involved, with some choices depending on external factors (e.g., the use of a FET-input follower for U_1 would be preferable for source impedances greater than about 50k).

Table 7.2 compares the specifications of op-amps you might choose for precision circuit design.

7.07 Amplifier output errors

As we discussed in Chapter 4, op-amps have some serious limitations associated with the output stage. Limited slew rate, output crossover distortion (see Section 2.15), and finite open-loop output impedance can all cause trouble, and they can cause precision circuits to display astoundingly large errors if not taken into account.

Slew rate: general considerations

As we mentioned in Section 4.11, an op-amp can swing its output voltage only at some maximum rate. This effect originates in the frequency-compensation circuitry of the op-amp, as we will explain in a bit more detail shortly. One consequence

TABLE 7.2. PRECISION OP-AMPS

Voltage				Current				Settling time, typ	Comments											
Offset		Drift		Bias		Offset														
Type	Mfg ^a	typ (μV)	max (μV)	typ (μV/°C)	max (μV/mo)	typ (nA)	max (nA)			e _n , typ (nV/√Hz)	i _n , typ (fA/√Hz)	PSRR min (dB)	Gain min (x1000)	Slew rate typ (V/μs)	f _T typ (MHz)	0.1% 0.01% (μs)				
BIPOLAR																				
OP-07A	PM+	10	25	0.2	0.6	0.2	0.7	2	0.3	2	10.3	9.6	320	120	100	300	0.17	0.6	-	classic prec op-amp
OP-08E	PM	70	150	0.5	2.5	-	0.8	2	0.05	0.2	22	20	150	130	104	80	0.12	0.8	-	improved 308
LM11	NS+	100	300	1	3	1	0.025	0.05	0.5pA	0.01	180	150	15	4	100	100	0.3	0.5	-	lowest bias bipolar
OP-12E	PM+	70	150	0.5	2.5	-	0.8	2	0.05	0.2	22	20	150	130	104	80	0.12	0.8	-	improved 312
OP-20B	PM	60	250	0.75	1.5	-	12	25	0.15	1.5	58	58	140	90	100	500	0.05	0.1	-	micropower
OP-21A	PM	40	100	0.5	1	-	50	100	0.6	4	21	21	380	210	104	1000	0.25	0.6	-	low power
OP-27E	PM+	10	25	0.2	0.6	0.5	10	40	7	35	3.5	3.0	1700	400	100	1000	2.8	8	-	low noise
OP-37E	PM+	10	25	0.2	0.6	0.5	10	40	7	35	3.5	3.0	1700	400	100	1000	17	63	-	low noise, decomp OP-27 (G>5)
OP-50E	PM	10	25	0.15	0.3	-	1	5	0.1	1	5.5	4.5	300	230	126	10M	3	25 ^b	-	hi curr, low noise, decomp (G>5)
OP-62E	PM	-	200	-	-	-	-	300	-	100	-	2.5	-	-	105	350	15	50	-	improved OP-07, dual=708
OP-77E	PM	10	25	0.1	0.3	0.3	1.2	2	0.1	1.5	10.3	9.6	320	120	110	5000	0.3	0.6	-	single supply; optional I _s = 30μA
OP-90E	PM	50	150	0.3	2	-	4	15	0.3	3	60	60	1600	700	104	700	0.01	0.02	-	low noise, ~OP-27
OP-97E	PM	10	25	0.2	0.6	0.3	0.03	0.1	0.03	0.1	17	14	20	6	114	300	0.2	0.9	-	low bias, imprvd 312; PM1012 ^c
MAX400M	MA	4	10	0.2	0.3	0.2	0.7	2	0.3	2	10.3	9.6	320	120	100	500	0.3	0.6	-	improved 358/324; sing supply ^s
LM607A	NS	15	25	0.2	0.3	0.2	1	2	0.5	2	9	6.5	320	120	100	5000	0.7	1.8	-	ultra low noise
AD707C	AD	5	15	0.03	0.1	0.2	0.5	1	0.1	1	10.3	9.6	320	120	120	8000	0.3	0.9	-	decomp 1007 (G>5), ~OP-37
LT1001A	LT	10	25	0.2	0.6	0.2	1	4	0.8	4	10.3	9.6	320	120	110	450	0.25	0.8	-	lowest non-chopper V _{os}
LT1006A	LT	20	50	0.2	1.3	0.4	9	15	0.12	0.5	23	22	70	30	106	1000	0.4	-	-	improved OP-07, dual=708
LT1007A	LT	10	25	0.2	0.6	0.2	10	35	7	30	2.8	2.5	1500	400	110	7000	2.5	8	-	single supply; optional I _s = 30μA
LT1012C	LT+	10	50	0.2	1.5	0.3	0.03	0.03	0.15	0.15	17	14	20	6	110	200	0.2	0.8	-	low noise, ~OP-27
LT1013A	LT	40	150	0.4	2	0.4	15	35	0.2	1.3	24	22	70	15	103	1500	0.4	0.8	-	low bias, imprvd 312; PM1012 ^c
LT1028A	LT	10	40	0.2	0.8	0.3	25	90	12	50	1	0.9	4700	1000	117	7000	15	75	-	improved 358/324; sing supply ^s
LT1037A	LT	10	25	0.2	0.6	0.2	10	35	7	30	2.8	2.5	1500	400	110	7000	15	60	-	ultra low noise
CA4077A	RA	4	10	0.1	0.3	0.2	0.3	2	0.1	1.5	10.3	9.6	320	120	110	2500	0.25	0.8	-	decomp 1007 (G>5), ~OP-37
HA5134	HA	25	250	-	5	-	-	-	25	-	7	-	2000	-	-	250	4	-	-	lowest non-chopper V _{os}
HA5135	HA	10	80	0.4	1.3	-	1	4	-	4	13	9	400	140	94	1000	0.8	2.5	11	quad, low noise
A5147A	HA	10	25	0.2	0.6	-	10	40	7	35	3.5	3.0	1700	400	80	1000	35	140	0.4	low noise, fast uncomm (G>10)

JFET

OPA101B	BB	50	250	3	5	-	6pA	0.01	1.5pA	4pA	25	8	1.4	1.4	86	60	7	20	2.5	10	low noise; decomp = OPA102
OPA111B	BB	50	250	0.5	1	-	0.5pA	1pA	0.3pA	0.7pA	30	7	0.4	0.4	100	1000	2	2	6	10	low noise, low bias
LFnnn	NS	1000	-	-	-	-	0.05	0.01	0.01	0.05	14	3.5	10	10	80	100	20	10	-	-	lowest noise JFET, no popcorn
LF455A	NS	75	180	3	4	-	7pA	0.05	3pA	0.02	100	12	10	10	86	200	5	3	-	4	456 and 457 faster
AD547L	AD	-	250	-	1	-	0.01	0.025	2pA	-	70	30	-	-	80	250	3	1	3.5	4.5	dual = AD642, 647
AD548C	AD	100	250	-	2	15	3pA	0.01	2pA	5pA	80	30	-	2	86	300	1.8	1	6	7	improved LF441; dual = AD648
AD711C	AD	100	250	2	3	15	0.015	0.025	5pA	0.01	45	18	-	10	86	200	20	4	0.9	1	improved LF411/2
LT1055A	LT	50	150	1.2	4	5	0.01	0.05	2pA	0.01	28	14	2	2	90	150	13	5	1.2	1.8	LT1056 is 20% faster
HA5170	HA	100	300	2	5	-	0.02	0.1	3pA	0.06	20	10	50	10	74	300	8	8	1	1.1	low noise

JFET, HIGH-SPEED

OP-44E	PM	30	750	4	10	5	0.08	0.2	4pA	0.04	38	13	-	7	90	500	120 ^h	16 ^h	0.2	-	decomp (G>3)
LF401A	NS	-	200	-	-	-	-	0.2	-	0.1	60	23	-	10	80	100	30	16	0.2	0.3	fast settle
OPA404B	BB	260	750	3	-	-	1pA	4pA	0.5pA	4pA	32	15	0.6	0.6	86	40	35	6.4	0.6	1.5	quad
OPA602C	BB	100	250	1	2	-	0.5pA	1pA	0.5pA	1pA	23	13	0.6	0.6	86	40	35	6.5	0.7	1	low bias, fast settle
OPA605K	BB	250	500	-	5	-	0.01	0.035	2pA	-	80	20	-	-	74	-	94	20	0.3	0.4	uncomp (G>50)
OPA606L	BB	100	500	3	5	-	5pA	0.01	0.4pA	5pA	30	13	1.3	1.3	90	100	35	13	1	2.1	improved LF356
AD744C	AD	100	250	2	3	15	0.03	0.05	0.01	0.02	45	18	-	10	92	250	75	13	0.4	0.5	low dist (3ppm); decomp (G>2)
AD845K	AD	100	250	1.5	3	-	0.25	0.4	0.015	0.05	80	25	-	100	98	200	100	16	0.3	0.3	fast settle
LT1022A	LT	80	250	1.3	5	-	0.01	0.05	2pA	0.01	28	14	2	2	88	150	26	8.5	0.8	1.8	

CHOPPER STABILIZED^a

MAX420E	MA	1	5	0.02	0.05	0.1	0.01	0.03	0.015	0.06	1.1 ^c	-	10	-	120	1000	0.5	0.5	-	-	±15V V _S ; 430 has C _{int}
MAX422E	MA	1	5	0.02	0.05	0.1	0.01	0.03	0.015	0.06	1.1 ^c	-	10	-	120	1000	0.13	0.13	-	-	±15V V _S ; 432 has C _{int}
LMC668A	NS	1	5	0.05	-	0.1	0.02	0.06	-	-	2 ^c	-	10	-	120	1000	2.5	1	-	-	
TSC900A	TS	-	5	0.02	0.05	-	-	0.05	0.5pA	-	4 ^c	-	-	-	120	1000	0.2	0.7	-	-	low power
TSC901	TS	7	15	0.05	0.15	-	0.03	0.05	0.05	0.1	5 ^c	-	-	-	120	1000	2	0.8	-	-	±15V supply; int caps
TSC911A	TS	5	15	0.05	0.15	-	-	0.07	5pA	0.02	11 ^c	-	-	-	112	600	2.5	1.5	-	-	int caps, noisy
TSC915	TS	-	10	0.01	0.1	-	0.03	0.1	0.05	0.1	0.8 ^c	-	-	-	120	1000	0.5	0.5	-	-	±15V supply
TSC918	TS	-	50	0.4	0.8	-	-	0.1	0.5pA	-	4 ^c	-	-	-	105	100	0.2	0.7	-	-	inexpensive
LTC1050	LT	0.5	5	0.01	0.05	0.05 ^d	0.01	0.03	0.02	0.06	1.6 ^c	-	2.2	-	125	300	4	2.5	-	-	int caps
LTC1052	LT	0.5	5	0.01	0.05	0.1 ^d	1pA	0.03	5pA	0.03	1.5 ^c	-	0.6	-	120	1000	4	1.2	-	-	improved 7652
ICL7650S	IL+	0.7	5	0.02	0.1	0.1	4pA	0.01	8pA	0.02	2 ^c	-	10	-	120	6000	2.5	2	-	-	improved 7650
ICL7652S	IL+	0.7	5	0.01	0.06	0.2	3pA	0.03	0.015	0.04	0.7 ^c	-	10	-	120	6000	1	0.5	-	-	improved 7652
TSC76HV52TS	-	10	-	0.3	-	-	0.03	0.1	0.05	0.1	0.8 ^c	-	-	-	120	1000	0.5	0.5	-	-	±15V 7652

(a) see footnotes to Table 4.1. (b) at G=50. (c) $\mu\text{V pp}$, 0.1-10Hz. (d) μV per square root month. (e) total supply=18V unless noted. (f) dual=1024. (g) quad=1014.

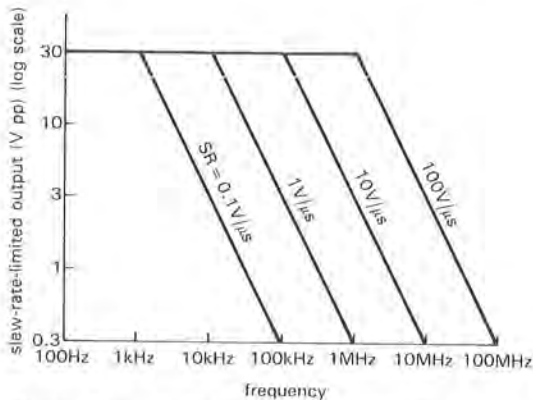


Figure 7.6. Maximum output swing versus frequency.

of finite slew rate is to limit the output swing at high frequencies, as we showed in Section 4.12 and illustrated in Figure 7.6:

$$V_{pp} = S/\pi f$$

A second consequence is best explained with the help of a graph of slew rate versus differential input signal (Fig. 7.7). The point to be made here is that a circuit that

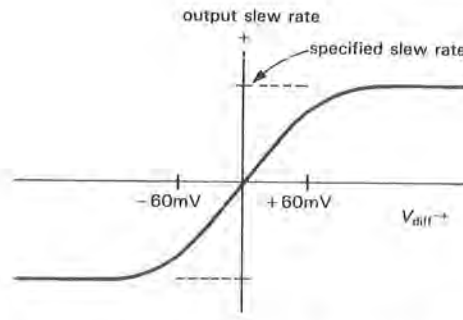


Figure 7.7. A substantial differential input voltage is required to produce full op-amp slew rate.

demands a substantial slew rate operates with a substantial voltage error across the op-amp's input terminals. This can be disastrous for a circuit that pretends to be highly precise.

Let's look at the innards of an op-amp in order to get some understanding of the origin of slew rate. The vast majority of op-amps can be summarized with the circuit

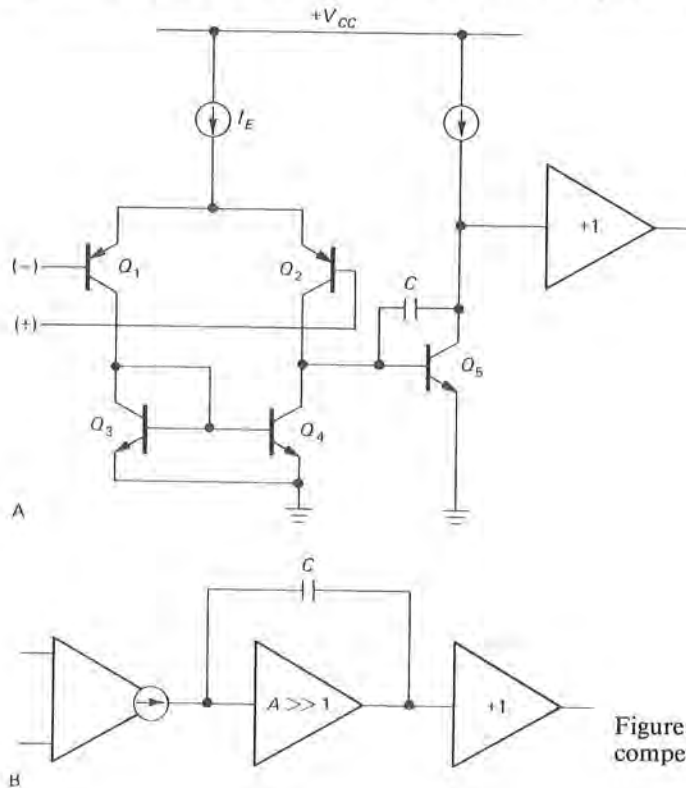


Figure 7.8. Typical op-amp internal compensation scheme.

shown in Figure 7.8. A differential input stage, loaded with a current mirror, drives a stage of large voltage gain with a compensation capacitor from output to input. The output stage is a unity-gain push-pull follower. The compensation capacitor C is chosen to bring the open-loop gain of the amplifier to unity before the phase shifts caused by the other amplifier stages have become significant. That is, C is chosen to put f_T , the unity-gain bandwidth, near the frequency of the next amplifier rolloff pole, as described in Section 4.34. The input stage has very high output impedance, and it looks like a current source to the next stage.

The op-amp is slew-rate-limited when the input signal drives one of the differential-stage transistors nearly to cutoff, driving the second stage with the total emitter current I_E of the differential pair. This occurs for a differential input voltage of about 60mV, at which point the ratio of currents in the differential stage is 10:1. At this point Q_5 is slewing its collector as rapidly as possible, with all of I_E going into charging C . Q_5 and C thus form an integrator, with a slew-rate-limited ramp as output. Read the accompanying section "Slew rate: a detailed look" to see how to derive an expression for the slew rate, knowing how bipolar transistors work.

□ SLEW RATE: A DETAILED LOOK

First, let us write an expression for the open-loop small-signal ac voltage gain, ignoring phase shifts:

$$A_V = g_m X_C = g_m / 2\pi f C$$

from which the unity-gain bandwidth product (the frequency at which $A_V = 1$) is

$$f_T = \frac{1}{2\pi} \frac{g_m}{C}$$

Now, the slew rate is determined by a current I_E charging a capacitance C :

$$S = \frac{dV}{dt} = \frac{I_E}{C}$$

For the usual case of a differential amplifier with no emitter resistors, g_m is related to I_E by

$$g_m = \frac{1}{r_e} = \frac{I_E}{2V_T} = \frac{I_E}{50\text{mV}}$$

By substituting this into the slew-rate formula, we find

$$S = 2V_T \frac{g_m}{C}$$

i.e., the slew rate is proportional to g_m/C , just the same as the unity-gain bandwidth! In fact,

$$S = 4\pi V_T f_T = 0.3 f_T$$

with f_T in MHz and S in V/ μ s. This is independent of the particular values of C , g_m , I_E , etc., and it gives a good estimate of slew rate (e.g., the classic 741, with $f_T \approx 1.5\text{MHz}$, has a slew rate of 0.5V/ μ s). It shows that an op-amp with greater gain-bandwidth product f_T will have a higher slew rate. You can't improve matters in a slow op-amp by merely increasing input-stage current I_E , because the increased gain (from increased g_m) then requires a correspondingly increased value of C for compensation. Adding gain anywhere else in the op-amp doesn't help either.

The preceding result shows that increasing f_T (by raising collector currents, using faster transistors, etc.) will increase the slew rate. A high f_T is, of course, always desirable, a fact not lost on the IC designer, who has already done the best he can with what's on the chip. However,

there is a way to get around the restriction that $S = 0.3f_T$. That result depended on the fact that the transconductance was determined by I_E (through $g_m = I_E/2V_T$). You can use simple tricks to raise I_E (and therefore the slew rate) while keeping f_T (and therefore compensation) fixed. The easiest is to add some emitter resistance to the input differential amplifier. Let's imagine we do something like that, causing I_E to increase by a factor m while holding g_m constant. Then, by going through the preceding derivation, you would find

$$S = 0.3mf_T$$

EXERCISE 7.1

Prove that such a trick does what we claim.

□ Increasing slew rate

Here, then, are some ways to obtain a high slew rate: (a) Use an op-amp with high f_T . (b) Increase f_T by using a smaller compensation capacitor; of course, this is possible only in applications where the closed-loop gain is greater than unity. (c) Reduce the input-stage transconductance g_m by adding emitter resistors; then reduce C or raise I_E proportionately. (d) Use a different input-stage circuit.

The third technique (reduced g_m) is used in many op-amps. As an example, the HA2605 and HA2505 op-amps are nearly identical, except for the inclusion of emitter resistors in the input stage of the HA2505. The emitter resistors increase the slew rate, at the expense of open-loop gain. The following data demonstrate this trade-off. FET op-amps, with their lower input-stage g_m , tend to have higher slew rates for the same reason.

	HA2605	HA2505
f_T	12MHz	12MHz
Slew rate	7V/ μ s	30V/ μ s
Open-loop gain	150,000	25,000

The fourth technique generally uses the method of "cross-coupled transconductance reduction," which involves having a second set of transistors available at the input stage, biding their time during small signal swings, but ready to help out with some extra current when needed. This has the advantage of improved noise and offset performance, at the expense of some complexity, as compared with the simple emitter resistor scheme. This technique is used in the Harris HA5141 and HA5151, Raytheon 4531, and Signetics 535 and 538 to boost the slew rate for large differential input signals. The resultant graph of slew rate versus input error signal is shown in Figure 7.9.

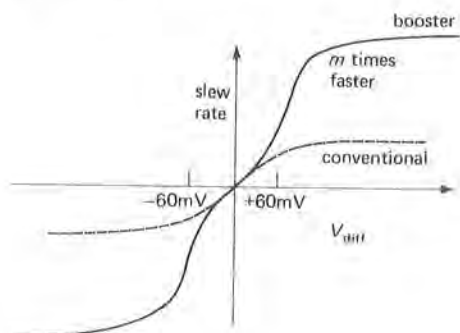
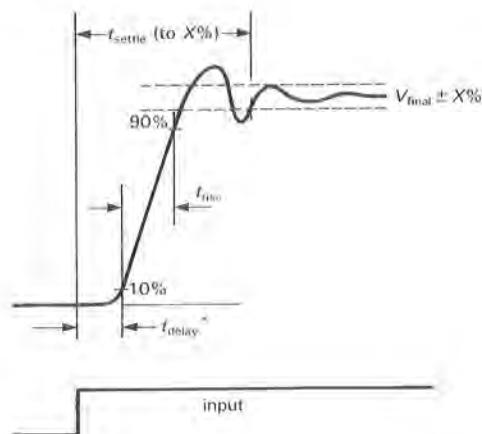


Figure 7.9

Bandwidth and settling time

Slew rate measures how rapidly the output voltage can change. The op-amp slew-rate specification usually assumes a large differential input voltage (60mV or more), which is realistic, since an op-amp whose output isn't where it's supposed to be will have its input driven hard by feedback, assuming a reasonable amount of loop gain. Of perhaps equal importance in high-speed precision applications is the time required for the output to get where it's going following an input change. This *settling-time* specification (the time required to get within the specified accuracy of the final value and stay there, Fig. 7.10) is always given for devices such as digital-to-analog converters, where precision is the name of the game, but it is not normally specified for op-amps.



* Sometimes defined to $V_{out} = \text{logic threshold}$, or $V_{out} = 0.5V_{total}$.

Figure 7.10. Settling time defined.

We can estimate op-amp settling time by considering first a different problem, namely what would happen to a perfect voltage step somewhere in a circuit if it were followed by a simple RC low-pass filter (Fig. 7.11). It is a simple exercise to show that the filtered waveform has

the settling times shown. This is actually an important result, since you often limit bandwidth with a filter to reduce noise (more on that later in the chapter). To extend this simple result to an op-amp, just remember that a compensated op-amp has a 6dB/octave rolloff, just like a low-pass filter. When connected for closed-loop gain G_{CL} , its “bandwidth” (the frequency at which the loop gain drops to unity) is approximately given by

$$f_{3dB} = f_T / G_{CL}$$

As a general result, a system of bandwidth B has response time $\tau = 1/2\pi B$; thus, the equivalent “time constant” of the op-amp is

$$\tau \approx G_{CL} / 2\pi f_T$$

The settling time is then roughly $5\tau - 10\tau$.

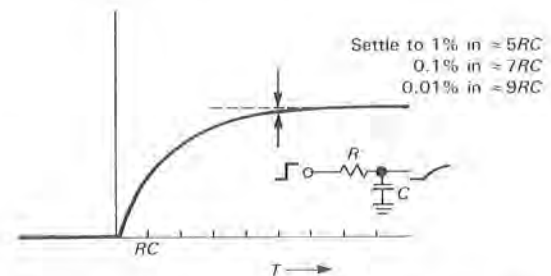


Figure 7.11. Settling time of an RC low-pass filter.

Let's try our prediction on a real case. The OP-44 from PMI is a precision fast-settling decompensated ($G_{CL} \geq 3$) op-amp, with an f_T of 23MHz (typ). Our simple formula then estimates the response time to be 21ns, which implies a settling time of $0.15\mu s$ (7τ) to 0.1%. This is in pretty good agreement with the actual value, which the data sheet gives as $0.2\mu s$ (typ) to 0.1%.

There are several points worth making: (a) Our simple model only gives us a lower bound for the actual settling time in a real

circuit; you should always check the slew-rate-limited rise time, which may dominate. (b) Even if slew rate is not a problem, the settling time may be much longer than our idealized “single-pole” model, depending on the op-amp’s compensation and phase margin. (c) The op-amp will settle more quickly if the frequency compensation scheme used gives a plot of open-loop phase shift versus frequency that is a nice straight line on a log-log graph (e.g., the OP-42, Fig. 7.12); op-amps with wiggles in the phase-shift graph are more likely to exhibit overshoot and ringing, as in the upper waveform shown in Figure 7.10. (d) A fast settling time to 1%, say, doesn’t necessarily guarantee a fast settling time to 0.01%, since there may be a long tail (Fig. 7.13). (e) There’s no substitute for an actual settling-time specification from the manufacturer.

Table 7.3 lists a selection of high-speed op-amps suitable for applications that demand high f_T , high slew rate, fast settling time, and reasonably low offset voltage.

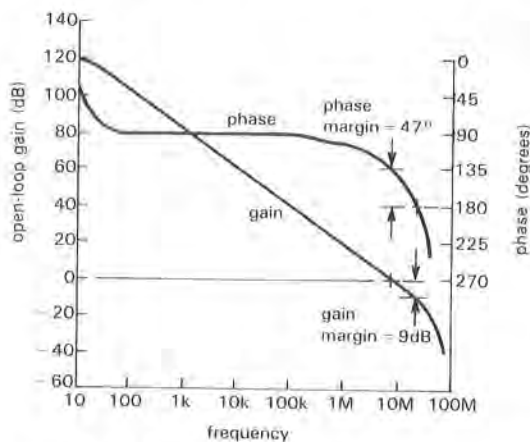


Figure 7.12. OP-42 gain and phase versus frequency.

Gain error

There’s one more error that arises from finite open-loop gain, namely an error in closed-loop gain owing to finite loop gain.

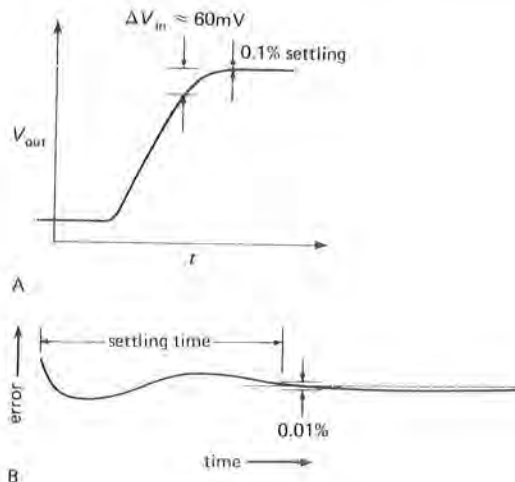


Figure 7.13. A. Slew rate decreases when input error approaches 60mV. B. Settling to high precision can be surprisingly lengthy.

We calculated in Chapter 3 the expression for closed-loop gain in a feedback amplifier, $G = A/(1 + AB)$, where A is the open-loop gain and B is the “gain” of the feedback network. You might think that the $A \geq 100\text{dB}$ of op-amp open-loop gain is plenty, but when you try to construct extremely precise circuits you are in for a surprise. From the preceding gain equation it is easy to show that the “gain error,” defined as

$$\delta_G = \text{gain error} \equiv \frac{G_{\text{ideal}} - G_{\text{actual}}}{G_{\text{ideal}}}$$

is just equal to $1/(1 + AB)$ and ranges from 0 for $A = \infty$ to 1 (100%) for $A = 0$.

EXERCISE 7.2

Derive the foregoing expression for gain error.

The resulting frequency-dependent gain error is far from negligible. For instance, a 411, with 106dB of low-frequency open-loop gain will have a gain error of 0.5% when configured for a closed-loop gain of 1000. Worse yet, the open-loop gain drops 6dB/octave above 20Hz, so our amplifier would have a gain error of 10% at

500Hz! Figure 7.14 plots gain error versus frequency, for closed-loop gains of 100 and 1000, for the OP-77, with its extraordinary 140dB of low-frequency gain. It should be obvious that you need plenty of gain and a high f_T to maintain accuracy at even moderate frequencies.

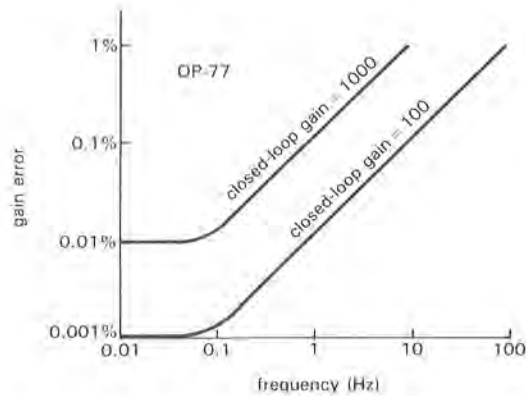


Figure 7.14. OP-77 gain error.

We plotted these curves using the graph of open-loop gain versus frequency given in the data sheet. Even if your op-amp data sheet provides a curve, it's best to work backward from the specified f_T and dc open-loop gain, figuring the open-loop gain at the frequency of interest, hence the gain error (as above) as a function of frequency. This procedure yields

$$\delta_G = \frac{1}{1 + Bf_T/f}$$

where B is, as usual, the gain of the feedback network. Of course, in some applications, such as filters, B may also depend on frequency.

EXERCISE 7.3

Derive the foregoing result for $\delta_G(f)$.

Crossover distortion and output impedance

Some op-amps use a simple push-pull output stage, without biasing the bases two

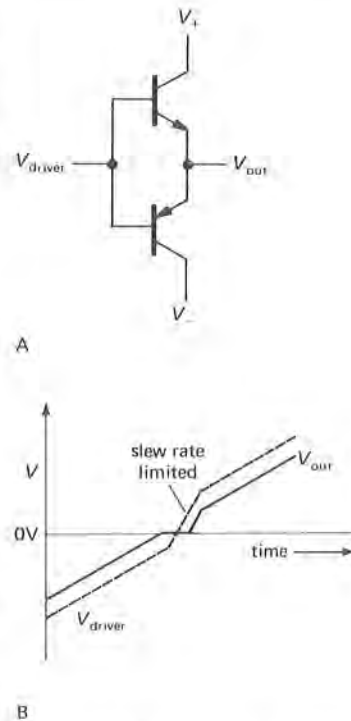


Figure 7.15. Crossover distortion in class B push-pull output stage.

diode drops apart, as we discussed in Section 2.15. This leads to class B distortion near zero output, since the driver stage has to slew the bases through $2V_{BE}$ as the output current passes through zero (Fig. 7.15). This crossover distortion can be substantial, particularly at higher frequencies where the loop gain is reduced. It is greatly reduced in op-amp designs that bias the output push-pull pair into slight conduction (class AB). The popular 741 is an example of the latter, whereas its predecessor, the 709, uses the simple class B output-stage biasing. The otherwise admirable 324 can exhibit large crossover distortion for this reason. The right choice of op-amp can have enormous impact on the performance of low-distortion audio amplifiers. Perhaps this problem has contributed to what the audiophiles refer to

TABLE 7.3. HIGH-SPEED PRECISION OP-AMPS

Type	Mfg ^a	Trt	Ext comp	Min gain	V _{os} max (mV)	ΔV _{os} max (μV/°C)	I _{os} max (nA)	I _b max (nA)	e _n @1kHz typ (nV/√Hz)	Input cap typ (pF)	Slew rate typ (V/μs)	f _T typ (MHz)	Settle, typ		R _{out} typ (Ω)	Over-shoot @G _{min} typ (%)	Phase margin @G _{min} (deg)	Diff'l input max (V)	Swing into load		Max output curr (mA)	Comments	
													0.1% (ns)	0.01% (ns)					±V _o (±V)	R _L (Ω)			
OP-37E	PM+	—	—	5	0.025	0.6	35	40	3.0	—	17	63	1000	—	70	10	71	0.7	11	600	5	low-ncise (decomp OP-27)	
OP-42E	PM	—	—	1	0.75	10	0.04	0.2	13	6	58	10	450	700	50	15	47	40	12	600	6	stable into 300pF	
OP-44E	PM	—	—	3	0.75	10	0.04	0.2	12	—	122	20	200	300?	—	25	53	40	—	—	7	decomp OP-42	
OP-62E	PM	—	—	1	0.2	—	100	300	2.5	—	15	50	—	—	—	—	—	5	12	600	—	—	
OP-63E	PM	—	—	1	0.75	—	100	300	7	—	50	50	—	—	—	—	—	5	12	600	—	—	
OP-64E	PM	—	—	5	0.75	—	100	300	7	—	200	200	—	—	—	—	—	5	12	600	—	—	
OP-65E	PM	—	—	1	2	—	2000	2.5μA	—	—	200	150	—	—	—	—	—	5	3.5	500	50	±5V supply	
CLC221	CL	—	—	1	1	15	—	10μA ^b	4	2.4	6500	—	15	18 ^c	8	12	—	—	—	—	30	hybrid	
CLC400	CL	—	—	1	5.5	40	—	25μA ^b	12	0.5	700	280	10	15	—	0	—	3.5	100	70	70	monolithic; transresistance	
CLC401	CL	—	—	7	6.5	50	—	35μA ^b	12	0.5	1200	2100	10	13	—	0	—	3.5	100	70	70	monolithic; transresistance	
LF401A	NS	—	—	1	0.2	10	0.1	0.2	20	7	70	16	200	340	50	—	55	32	12	600	12	—	
OPA602C	BB	—	—	1	0.25	2	0.001	0.001	13	3	35	6.5	700	1000	80	20	40	36	10	600	4	low-bias	
AD711C	AD	—	—	1	0.25	5	0.01	0.025	18	6	20	4	1000	—	2	10	—	20	10	600	2.8	inexpensive	
MS738B	KE	—	—	2	0.075	1	20	40	3.8	3	3500	1500	30	200	2	5	—	12	12	100	120	—	
MS739	KE	—	—	1	0.025 ^d	0.3 ^t	0.03 ^t	0.08 ^t	—	—	5500	—	15	30 ^c	—	—	—	—	12	100	120	—	
AD744C	AD	—	—	2	0.25	3	0.02	0.05	18	5.5	75	13	400	500	—	4	70	36	12	600	4	ultra low dist (3ppm) ^f	
AD840K	AD	—	—	10	0.5	5	200	5μA	3	1	400	400	80	110	15	20	—	10	500	50	50	decomp 841	
AD841K	AD	—	—	1	1	20	200	5μA	13	1	300	40	80	110	5	20	—	10	500	12	12	vert PNP, decomp avail	
AD842K	AD	—	—	2	1	10	200	5μA	9	1	375	80	80	110	5	20	—	10	500	100	100	decomp 841	
AD845K	AD	—	—	1	0.25	3	0.05	0.4	25	4	100	16	250	300	—	—	—	20	12.5	500	10.2	—	
AD846K	AD	—	—	2	0.2	2	—	150 ^b	1.3	2	450	40	80	110	16	20	—	18	10	500	7	low-noise	
AD847J	AD	—	—	1	1	30	15 ^t	5μA	15 ^b	1.5	300	50	80	—	—	—	50	6	10	500	5.6	vert PNP, decomp avail	
AD848J	AD	—	—	1	1	10	15 ^t	5μA	4 ^d	1.5	300	250	80	—	—	—	50	6	2.5	150	20	vert PNP	
LT1028A	LT	—	—	2	0.04	0.8	50	90	0.85	5	15	75	—	—	80	5	50	0.7	12	600	10.5	ultra low noise	
LT1055A	LT	—	—	1	0.15	4	0.01	0.05	14	4	13	5	1200	1800	60	10	—	40	10	600	4	LT1056 is faster	
1435	TP	—	—	2	5	25	300 ^t	20μA	16	2	300	1000	40	70	—	1	25	4	7	500	30	hybrid	
LH4105C	NS	—	—	1	0.5	20 ^t	0.4	0.6	—	—	40	18	—	500	—	—	—	30	10	100	25	no current limit	
HA5147A	HA	—	—	10	0.025	0.6	35	40	3.0	—	35	140	400	—	70	20	—	0.7	11	600	4	low-noise (decomp OP-27)	
AD9611B	AD	—	—	1	3	20	—	5	1 ^e	3	1900	280	13	—	0.03	4	—	—	3	100	50	50	cur fdbk; no protec; hybrid

[a] see notes to Table 4.1.

[b] current-sensing inv input; bias current shown is for non-inv input only.

[c] to 0.02%.

[d] at 10kHz.

[e] 5MHz to 280MHz.

[f] typical

(a) see notes to Table 4.1.

(b) current-sensing inv input; bias current shown is for non-inv input only.

(c) to 0.02%.

(d) at 10kHz.

(e) 5MHz to 280MHz.

(f) typical

as “transistor sound.” Some modern op-amps, particularly those intended for audio applications, are designed to produce extremely low crossover distortion. Examples are the LT1028, the LT1037, and the LM833. The LM833, for example, has less than 0.002% distortion over the full audio band of 20Hz–20kHz. (That’s the claim, anyway; we may be overly gullible!) These amplifiers all have very low noise voltage, as well; in fact, the LT1028 is currently the world noise-voltage champion, with $e_n = 1.7\text{nV}/\sqrt{\text{Hz}}$ (max) at 10Hz.

The open-loop output impedance of an op-amp is highest near zero output voltage, because the output transistors are operating at their lowest current. The output impedance also rises at high frequency as the transistor gain drops off, and it may rise slightly at very low frequencies due to thermal feedback on the chip.

It is easy to neglect the effects of finite open-loop output impedance, thinking that feedback will cure everything. But when you consider that some op-amps have open-loop output impedances of a few hundred ohms, it becomes clear that the effects may not be negligible, especially at low to moderate loop gains. Figure 7.16 shows some typical graphs of op-amp output impedance, both with and without feedback.

Driving capacitive loads

The finite open-loop output impedance of op-amps leads to serious difficulties when you attempt to drive a capacitive load, owing to lagging phase shifts produced by the output impedance in combination with the load capacitance to ground. These can lead to feedback instabilities if the 3dB frequency is low enough, since it adds to the 90° phase shift already present with frequency compensation. As an example, imagine driving a hundred feet of coaxial cable from an op-amp with 200 ohms output impedance. The unterminated coax line acts like a 3000pF capacitor,

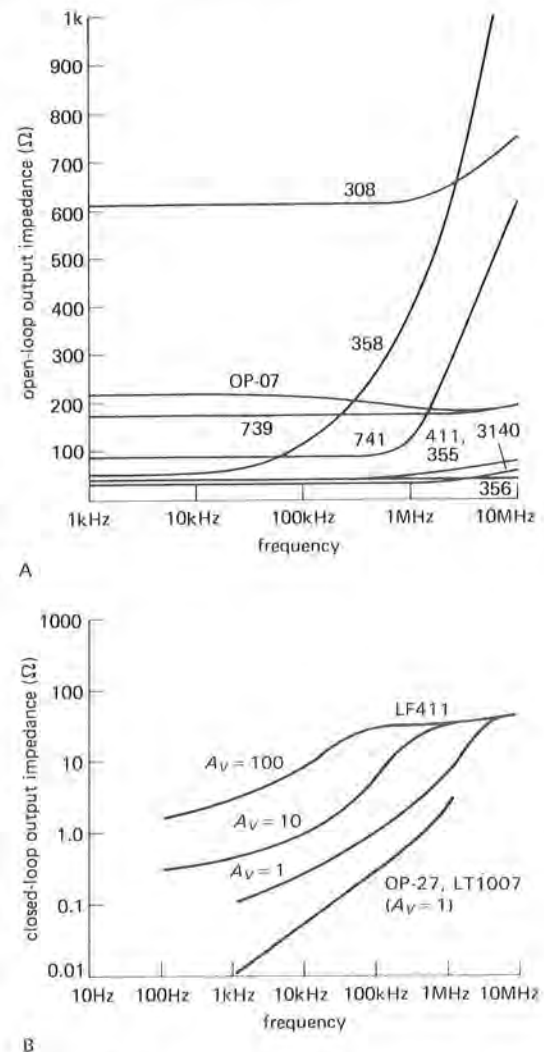


Figure 7.16. A. Measured open-loop output impedance for some popular op-amps. B. Closed-loop output impedance for the 411 and OP-27 op-amps.

generating a low-pass RC with a 3dB point of 270kHz. This is well below the unity-gain frequency of a typical op-amp, so oscillations are likely at high loop gain (a follower, for example).

There are a couple of solutions to this problem. One is to add a series resistor, taking feedback at high frequencies from the op-amp output and feedback at low frequencies and dc from the cable (Fig. 7.17).

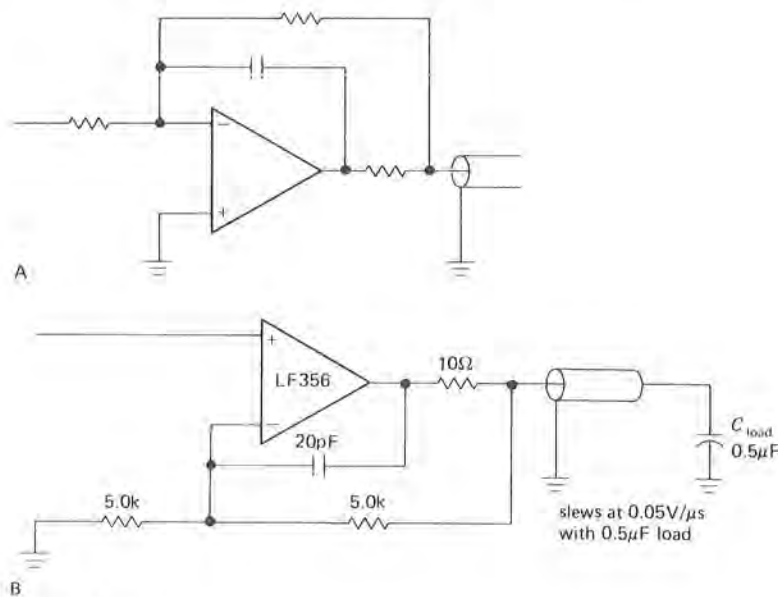


Figure 7.17

The parts values shown in the second circuit are specific for that op-amp and circuit configuration, and they give an idea of how large a capacitance can be driven. Of course, this technique degrades the high-frequency performance, since feedback isn't operative at high frequencies on the signal at the cable.

Unity-gain power buffers

If this technique of split feedback paths is unacceptable, the best thing to do is add a unity-gain high-current buffer inside the loop (Fig. 7.18). The devices listed have

voltage gain near unity and low output impedance, and they can supply up to 250mA output current. They have no significant phase shifts up to the unity-gain frequency (f_T) of most op-amps, and they can be included in the feedback loop without any additional frequency compensation. Table 7.4 presents a brief listing of buffer amplifiers. These "power boosters" can, of course, be used for loads that require high current, regardless of whether or not there are problems with capacitance. Unfortunately, most buffer amplifiers do not contain either internal current limiting or thermal shutdown circuits and must therefore

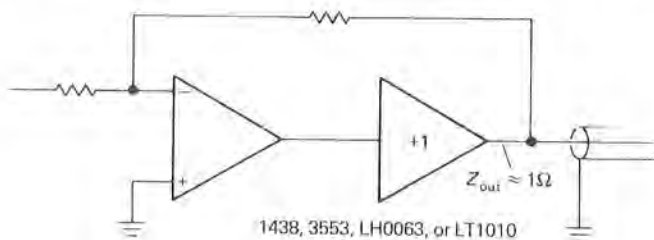


Figure 7.18

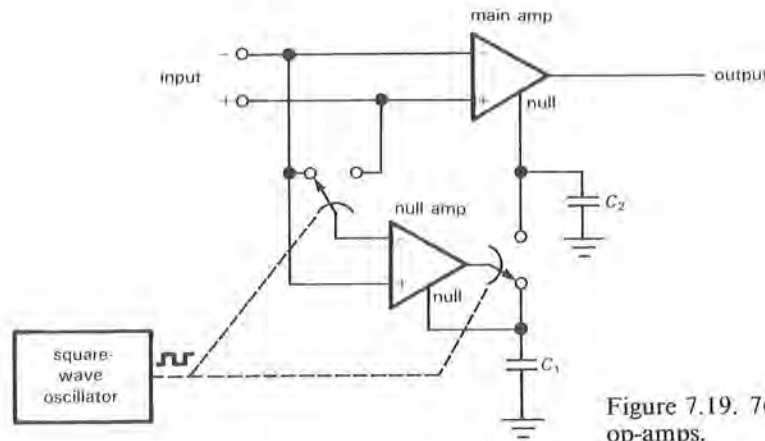


Figure 7.19. 7650-type chopper-stabilized op-amps.

be used carefully. The exceptional devices that include on-chip protection are noted in Table 7.4, for example the LT1010.

Note also that the preceding example would be changed if the cable were terminated in its characteristic impedance. In that case it would look like a pure resistance, somewhere in the range of 50 to 100 ohms, depending on the type of cable. In such a case a buffer would be mandatory, with $\pm 200\text{mA}$ drive capability in order to drive ± 10 volt signals into the 50 ohm load impedance. This subject is discussed in greater detail in Section 13.09.

The preceding circuit example does not suffer from any op-amp output-related errors, since it operates essentially at dc.

7.08 Auto-zeroing (chopper-stabilized) amplifiers

Even the best of precision low-offset op-amps cannot match the stunning V_{os} performance of the so-called “chopper-stabilized” or “auto-zero” op-amps. Ironically, these interesting amplifiers are built with CMOS, otherwise famous for its mediocrity when it comes to offset voltage or drift. The trick here is to put a second *nulling* op-amp on the chip, along with some MOS analog switches and offset-error storage capacitors (Fig. 7.19). The main op-amp functions as a conventional

imperfect amplifier. The nulling op-amp’s job is to monitor the input offset of the main amplifier, adjusting a slow correction signal as needed in an attempt to bring the input offset exactly to zero. Since the nulling amplifier has an offset error of its own, there is an alternating cycle of operation in which the nulling amplifier corrects its own offset voltage.

Thus, the auto-zeroing cycle goes like this: (a) Disconnect nulling amplifier from input, short its inputs together, and couple its output back to C_1 , the holding capacitor for its correction signal; the nulling amplifier now has zero offset. (b) Now connect nulling amplifier across input, and couple its output to C_2 , the holding capacitor for the main amplifier’s correction signal; the main amplifier now has zero offset. The MOS analog switches are controlled by an on-board oscillator, typically at a rate of a few hundred hertz. The error-voltage holding capacitors are typically $0.1\mu\text{F}$ and in most cases must be supplied externally; LTC, Maxim, and Teledyne make some convenient auto-zero amplifiers with discrete capacitors encapsulated right into the IC package.

Auto-zero op-amps do best what they are optimized for, namely delivering V_{os} values (and tempcos) five times better than the best precision bipolar op-amp (see Table 7.2). What’s more, they do this while

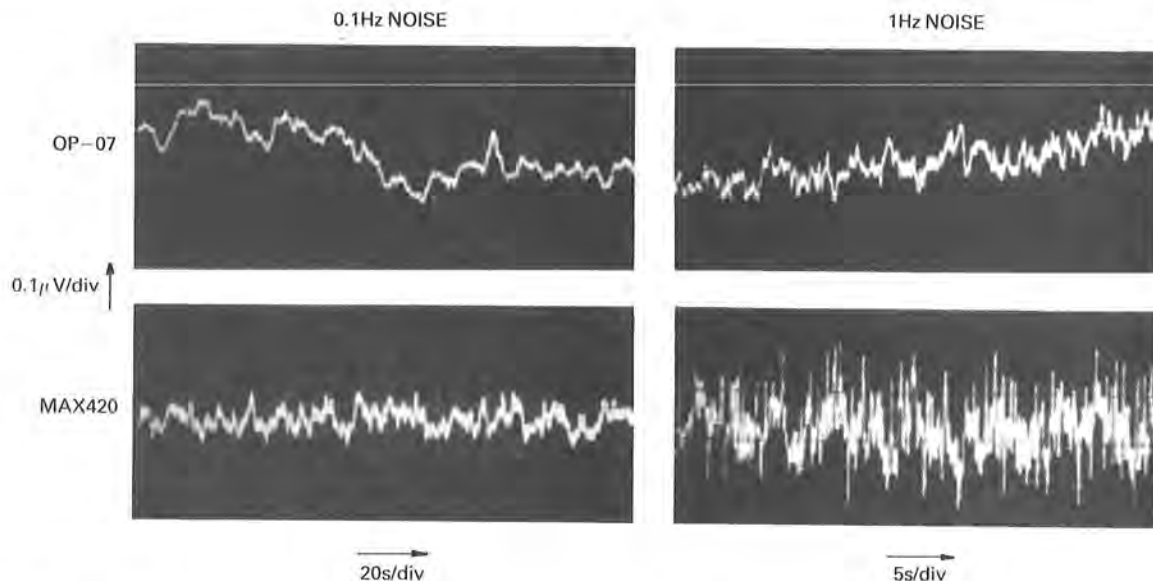


Figure 7.20. At very low frequencies a chopper-stabilized op-amp has lower noise than a conventional low-noise op-amp.

A. dc to 0.1Hz

B. dc to 1Hz (Courtesy of Maxim Integrated Products, Inc.)

delivering full op-amp speed and bandwidth, unlike earlier synchronous amplifiers that were also called “chopper amplifiers,” but which had bandwidth limited to a fraction of the chopping clock frequency (see below).

That’s the good news. The bad news is that auto-zero amplifiers have a number of diseases that you must watch out for. First of all, being CMOS devices, most of them have a severely limited supply voltage (typically 15V total supply) and thus cannot run from conventional ± 15 volt supplies. The Maxim MAX430/2 and Teledyne TSC915 and TSC76HV52 “high-voltage” auto-zero op-amps are exceptions and will operate from ± 15 volt supplies. Second, most auto-zero op-amps require external capacitors (exceptions: LTC1050, Maxim MAX430/2, Teledyne TSC911/13/14). A third problem with many auto-zero amplifiers (particularly those with limited supply voltage) is the rather restricted common-mode input range; for example, the popular ICL7650

has a guaranteed common-mode input range of -5 to $+1.5$ volts, when running from its usual ± 5 volt supplies (for the improved ICL7652 the range is -4.3 V to $+3.5$ V; that’s a wider range, but it doesn’t include the negative rail, so you can’t use it as a “single-supply” op-amp). The high-voltage amplifiers are much better – for example, the MAX432 has a guaranteed common-mode range of -15 to $+12$ volts, when running from ± 15 volt supplies. The op-amp table (Table 4.1) shows which chopper amps have common-mode range to the negative rail; although the popular ICL7652 doesn’t, improved versions from LTC (LTC1052) and Maxim (ICL7652B) do, permitting convenient single-supply operation.

A fourth drawback is the tendency of these CMOS op-amps to have poor output sourcing capability, sometimes as little as 1–2mA in the sourcing (positive-output) direction. The otherwise admirable MAX432 can only source 0.5mA! Fifth on the list of drawbacks, but often

first in importance, is the problem of clock-induced noise. This is caused by charge coupling from the MOS switches (see Section 3.12) and can cause wicked spikes at the output. The specifications are often misleading here, because it is conventional to quote input-referred noise with $R_S = 100\ \Omega$ and also to give the specification only for very low frequencies; for example, a typical input-referred noise voltage is $0.2\ \mu\text{V}$ (dc to 1 Hz, with $R_S = 100\ \Omega$). However, with zero input signal the output waveform might consist of a train of $5\ \mu\text{s}$ -wide 15 mV spikes of alternating polarity! In low-frequency applications you can (and should) RC-filter the output to a bandwidth of a few hundred hertz, which will make these spikes disappear. This spiky noise is also of no importance in integrating applications (e.g., integrating A/D converters, see Section 9.21) or in applications where the output is intrinsically slow (e.g., a thermocouple circuit with a meter at the output). In fact, if you only want very slow output response, and therefore low-pass-filter the output to extremely low frequencies (below 1 Hz), a chopper amplifier will actually have *less* noise than a conventional low-noise op-amp; see Figure 7.20.

A final problem with auto-zero amplifiers is their disastrous saturation characteristic. What happens is this: The auto-zeroing circuit, in attempting to bring the input difference voltage to zero, implicitly assumes there is overall feedback operating. If the amplifier's output saturates (or if there is no external circuit to provide feedback), there will be a large differential input voltage, which the nulling amplifier sees as an input offset error; it therefore blindly generates a large correction voltage that charges up the correction capacitors to a large voltage before the nulling amplifier itself finally saturates. Recovery is incredibly slow – up to a *second*! The “cure” is to sense when the output is approaching saturation, and clamp the input

to prevent it. Most auto-zeroing op-amps provide a “clamp” output for this purpose, which you tie back to the inverting input to prevent saturation. You can prevent saturation in chopper amplifiers without a “clamp” pin (and in ordinary op-amps, as well) by bridging the feedback network with a bidirectional zener (two zeners in series), which clamps the output at the zener voltage, rather than letting it limit at the supply rail; this works best in the inverting configuration.

□ Chopper miscellany

□ *ac-coupled “chopper amp.”* When considering auto-zeroing chopper amplifiers, be sure you don't confuse this technique with another “chopper” technique, namely the traditional low-bandwidth chopper amplifier in which a small dc signal is converted to ac (“chopped”) at a known frequency, amplified in ac-coupled amplifiers, then finally demodulated by multiplying with the same waveform used to chop the signal initially (Fig. 7.21). This scheme is quite different from the full-bandwidth auto-zeroing technique we've been considering, in that it rolls off at signal frequencies approaching the clock frequency, typically just a few hundred hertz. You sometimes see it used in chart recorders and other low-frequency instrumentation.

□ *Thermal offsets.* When you build dc amplifiers with submicrovolt offset voltages, you should be aware of *thermal offsets*, which are little thermally driven batteries produced by the junction of dissimilar metals (see Section 15.01). You get a Seebeck-effect “thermal emf” when you have a pair of such junctions at different temperatures. In practice you usually have joints between wires with different plating; a thermal gradient, or even a little draft, can easily produce thermal voltages of a few microvolts. Even similar wires from different manufacturers can produce thermal emf's of $0.2\ \mu\text{V}/^\circ\text{C}$, four times the drift

TABLE 7.4. FAST BUFFERS

Small signal				Supply voltages		Large signal				Comments		
Type	Mfg ^a	Rolloff freq		min		Slew rate (V/ μ s)	Maximum output current ($\pm$ mA)	Output swing			V_{os} max (mV)	
		-3dB (MHz)	-40' (MHz)	Z _{out} (Ω)	min ($\pm$ V)			max ($\pm$ V)	V _{out} ($\pm$ V)			R _{load} (Ω)
LT1010	LT	40	15	7	2.5	20	200	150	12	80	150	thermal limit; monolithic
LH0002	NS	50	60	6	6	22	200	100	10	50	30	
LH4001	NS	50	-	6	5	22	125	200	10	50	50	10-pin DIP
LM6321	NS	50	40	5	5	16	800	300	10	50	50	mini-DIP; therm lim; monolithic
AH0010	OE	60	-	20	6	18	1500	100	10	100	20	mini-DIP; alias 9910
BUF03	PM	65	20	2	6	18	250	70	10	150	6	monolithic
EL2001	EL	70	-	-	5	15	500	100	-	-	-	mini-DIP; monolithic
LH0033	NS+	100	80	6	5	20	1400	100	10	50	10	also EL2033, and others
1490	TP	100	-	20	12	18	500	100	-	-	20	FET input
HA5002	HA	110	-	3	-	20	1300	200	11	100	20	monolithic
HOS100	AD	125	-	8	5	20	1500	100	10	100	10	
MAX460	MA	140	65	4	5	20	1500	100	10	100	5	monolithic
LH4004	NS	140	-	-	4	15	1500	-	10	50	15	FET input; ext feedback
EL2005	EL	140	60	4	5	15	1500	100	10	100	5	FET input, precision
EL2002	EL	180	-	-	5	15	1000	100	-	-	-	mini-DIP; monolithic
LH0063	NS+	200	30	1	5	20	4000	250	10	50	25	"damn fast" buffer
MSK330	KE	200	-	2	-	18	3000	200	13	100	25	video
LH4002	NS	200	150	6	4	6	1250	60	2.2	50	50	
9911	OE	200	-	6	11	18	1000	500	10	20	20	FET input
9963	OE	200	-	3	6	18	3000	200	10	50	50	
1359	TP	250	60	5	12 ^a	18	1300	100	10	100	15	
LH4003	NS	250	-	-	5	8	1200	100	3	50	15	video; ext feedback
HA5033	HA	250	80	5	5	20	1300	100	10	100	15	mini-DIP; monolithic; also AH001
OPA633	BB	275	150	5	5	16	2500	100	11	50	15	monolithic
3553	BB	300	60	1	5	20	2000	200	10	50	50 ¹	insulated metal case
MP2004	MP	350	280	4	5	20	2500	100	10	100	10	FET input; also EL2004
LH4006	NS	350	-	-	4	8	1200	100	3	50	15	video; ext feedback
EL2031	EL	500	-	-	-	-	5000	100	-	-	-	FET input
CLC110	CL	730	200	2	5?	7	800	70	4	100	8	monolithic

(a) see footnotes to Table 4.1. (b) nominal. (t) typical.

(a) see footnotes to Table 4.1. (b) nominal. (f) typical.

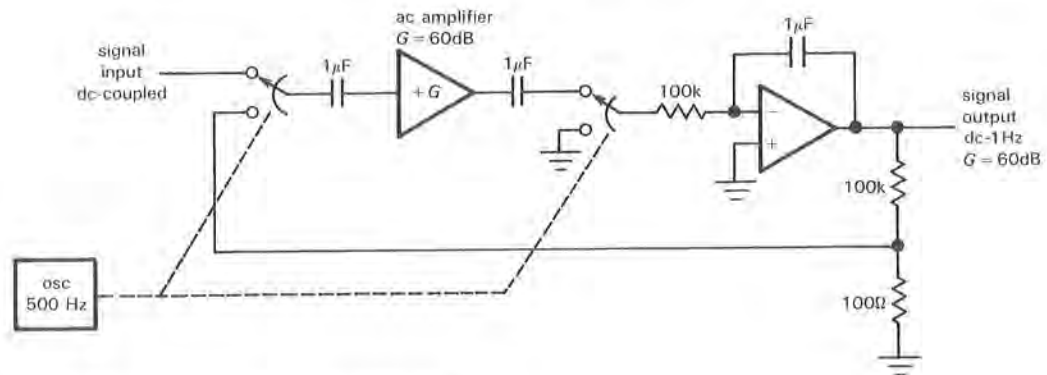


Figure 7.21. An ac-coupled chopper amplifier.

spec of a MAX432! The best approach is to strive for symmetrical wiring and component layouts, and then avoid drafts and gradients.

- **External auto-zero.** National makes a nice “auto-zero” chip (the LMC669) that can be used as an outboard nulling amplifier to make any op-amp of your choosing into an auto-zeroing amplifier (Fig. 7.22). It works most naturally with the inverting configuration, as shown, generating an error voltage to the noninverting input to maintain zero input offset. It doesn’t do as well as the dedicated auto-zero amplifiers we’ve been considering: The V_{os} specification is $5\mu\text{V}$ (typ), $25\mu\text{V}$ (max). However, it does let you use the auto-zero technique with any op-amp. You might, for instance, use it to zero an imprecise but high-power or high-speed op-amp. The circuits shown are good examples. The LM675 is a nice high-power op-amp (3A output current, with sophisticated on-chip safe-operating-area and thermal protection), but with a maximum offset voltage of 10mV. The auto-zero reduces that by nearly a factor of 1000. Likewise, the LM6364 is a fast op-amp ($f_T = 175\text{MHz}$, $\text{SR} = 350\text{V}/\mu\text{s}$) with maximum offset voltage of 9mV, here reduced by a factor of 400. Note the RC filter

components at both input and output of the auto-zero: These are necessary to suppress chopper noise in the (slow) correction loop, when this technique is used with small signals and low-noise parts like the LM6364 ($8\text{nV}/\sqrt{\text{Hz}}$).

- **Instrumentation amplifier.** Another “chopper” technique in use is the so-called “commutating auto-zeroing” (or CAZ) amplifier, originated by Intersil. In this technique, typified by the ICL7605 flying-capacitor instrumentation amplifier, MOSFET switches enable you to store the differential input signal across a capacitor, then amplify it with a single-ended chopper-stabilized amplifier (Fig. 7.23). You get charge-coupled spikes at the clock rate, just as with the standard auto-zeroing amplifier, which puts the same sort of limitations on the CAZ technique as we saw earlier. Although we raved about CAZ amplifiers in our first edition (“...stands a good chance of revolutionizing precision op-amp and instrumentation amplifier technology”), they’ve been finessed by the better auto-zeroing technique in which the signal always passes through a single amplifier.

However, in fairness to the CAZ-amp, we should point out that the flying-capacitor technique used in the 7605 has

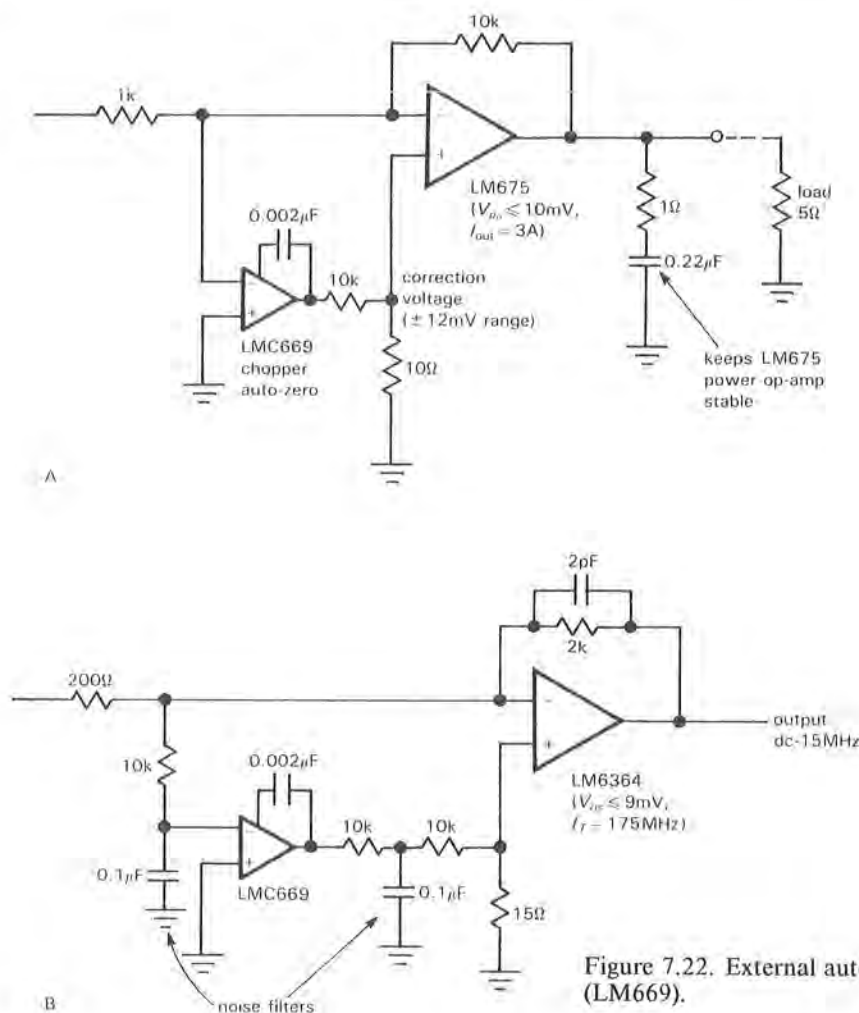


Figure 7.22. External auto-zeroing IC (LM669).

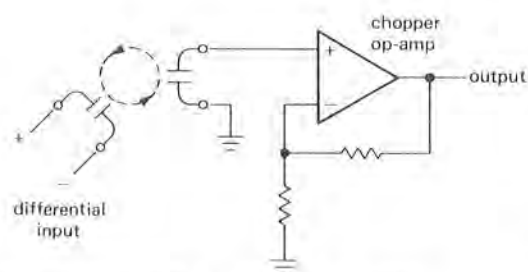


Figure 7.23. ICL7605 flying-capacitor differential amplifier with high CMRR.

some unique advantages, including input common-mode operation 0.3 volt beyond both supply rails, 100dB CMRR (min),

even at unity gain, and the lowest offset voltage of any monolithic amplifier. If you use these amplifiers, however, don't forget the required output noise filter, the limited supply voltage ($\pm 8\text{V}$ max), and the requirement of a high-impedance load, since the output impedance rises periodically at the clock rate.

The LTC1043 flying-capacitor building block lets you make your own high-CMRR differential amplifier. Instrumentation amplifiers are discussed in detail in the next section. The precision op-amp table (Table 7.2) includes most of the currently available auto-zeroing op-amps.

DIFFERENTIAL AND INSTRUMENTATION AMPLIFIERS

The term *instrumentation amplifier* is used to denote a high-gain dc-coupled differential amplifier with single-ended output, high input impedance, and high CMRR. They are used to amplify small differential signals coming from transducers in which there may be a large common-mode signal or level.

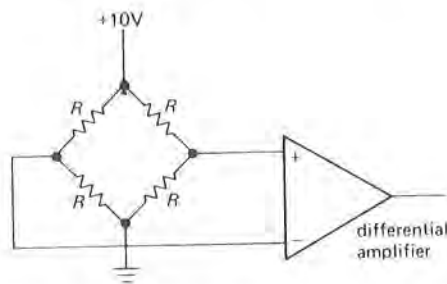


Figure 7.24. Strain gauge with amplifier.

An example of such a transducer is a strain gauge, a bridge arrangement of resistors that converts strain (elongation) of the material to which it is attached into resistance changes (see Section 15.03); the net result is a small change in differential output voltage when driven by a fixed dc bias voltage (Fig. 7.24). The resistors all have roughly the same resistance, typically 350 ohms, but they are subjected to differing strains. The full-scale sensitivity is typically 2mV per volt, so that the full-scale output is 20mV for 10 volts dc excitation. This small differential output voltage proportional to strain rides on a 5 volt dc level. The differential amplifier must have extremely good CMRR in order to amplify the millivolt differential signals while rejecting variations in the ~ 5 volt common-mode signal. For example, suppose that a maximum error of 0.1% is desired. Since 0.1% of full scale is 0.02mV, riding on 5000mV, the CMRR would have to exceed 250,000 to 1, i.e., about 108dB.

The tricks involved in making good instrumentation amplifiers and, more generally, high-gain differential amplifiers are similar to the techniques just discussed. Bias current, offsets, and CMRR errors are all important. Let's begin by discussing the design of differential amplifiers for noncritical applications first, working up to the most demanding instrumentation requirements and their circuit solutions.

7.09 Differencing amplifier

Figure 7.25 shows a typical circuit situation requiring only modest common-mode rejection. This is a current-sensing circuit used as part of a constant-current power supply to generate a constant current in the load. The drop across the precision 4-wire 0.01 ohm power resistor is proportional to load current. Even though one side of R_5 is connected to ground, it would be unwise to use a single-ended amplifier, since connection resistances of a milliohm would contribute 10% error! A differential amplifier is obviously required, but it need not have particularly good CMRR, since only very small common-mode signals are expected.

The op-amp is connected in the standard differencing amplifier configuration, as discussed in Section 4.09. R_1 , R_2 , and R_5 are precision wire-wound types for extreme stability of gain, whereas R_3 and R_4 , which set CMRR, can be mere 1% metal-film types. The overall circuit thus has a gain accuracy approaching that of the current-sensing resistor and a CMRR of about 40dB.

Precision differential amplifier

For applications such as strain gauges, thermocouples, and the like, 40dB of common-mode rejection is totally inadequate, and figures more like 100dB to 120dB are often needed. In the preceding example of the strain gauge, for instance, you might have a full-scale differential (unbalance) signal

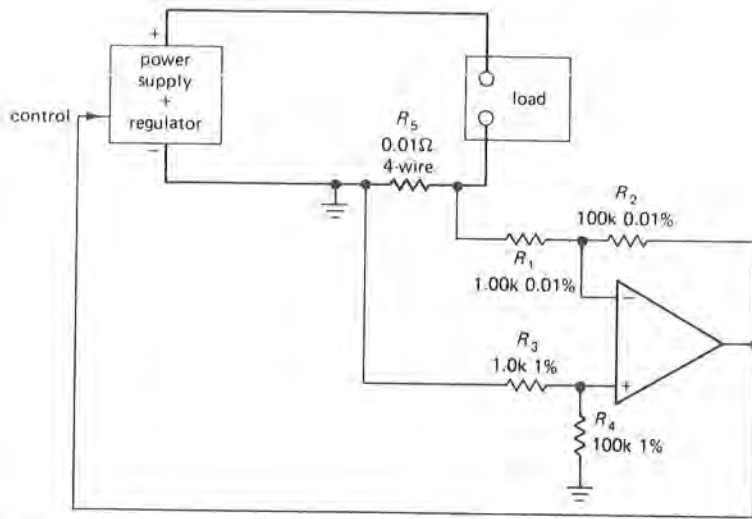


Figure 7.25. Current regulator.

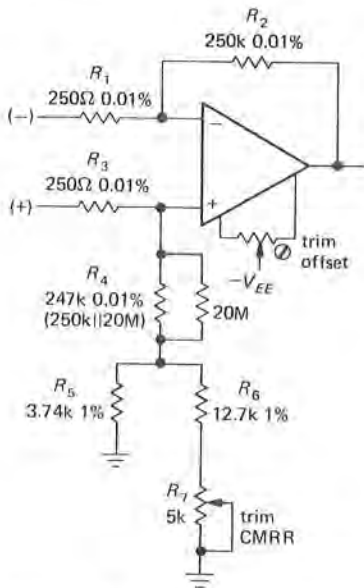


Figure 7.26

of 2mV per volt. If you want accuracy of 0.05%, you need a common-mode rejection of 114dB, minimum. (Note that this requirement can be relaxed considerably in the special case that the amplifier is zeroed with the common-mode voltage present, as might be done in a laboratory situation.)

The obvious first approach to improved CMRR is to beef up the resistor precision in the differencing circuit (Fig. 7.26). The resistor values are chosen to keep the large feedback resistors within the range of available precision wire-wound resistors. With 0.01% resistors, the common-mode rejection is in the range of 80dB (68dB worst case), assuming the op-amp has high CMRR. It takes only one trimmer to null the common-mode sensitivity, as shown. With the values shown, you can trim out an accumulated error up to 0.05%, i.e., a bit more than the worst-case resistor error. The fancy network shown is used because small-value trimmer resistors tend to be somewhat unstable with time and are best avoided.

A point about ac common-mode rejection: With good op-amps and careful trimming, you can achieve 100dB or better CMRR at dc. However, the wire-wound resistors you need for the best stability have some inductance, causing degradation of CMRR with frequency. Noninductive wire-wound resistors (Aryton-Perry type) are available to reduce this effect, which is common to all the circuits we

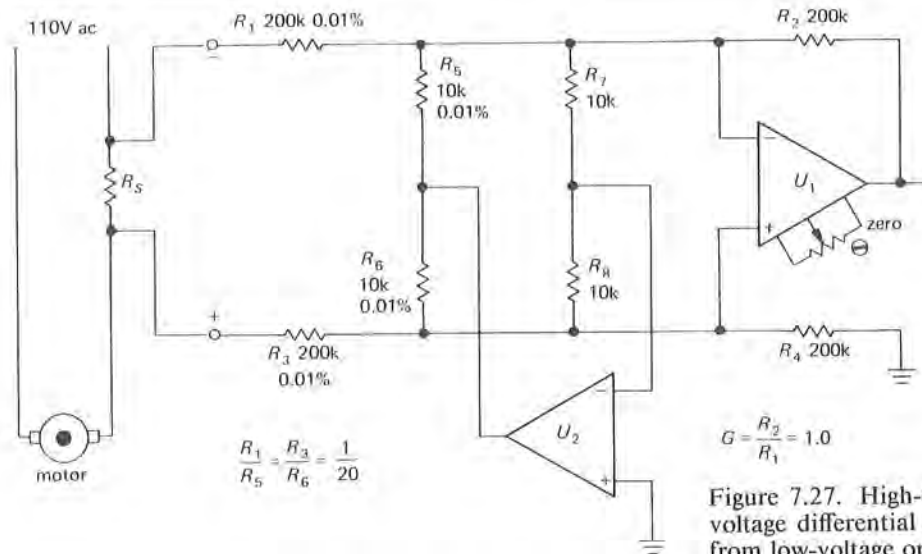


Figure 7.27. High-common-mode-voltage differential amplifier made from low-voltage op-amps.

will be talking about. Note also that it is necessary to balance the circuit capacitances to achieve good CMRR at high frequencies. This may require careful mirror-image placement of components.

Burr-Brown offers a series of precision differential amplifiers, complete with matched resistors, in a nice mini-DIP (8-pin) package. The INA105 is unity-gain ($\pm 0.01\%$ maximum gain error), with input impedance of $25k\Omega$, and the INA106 has a gain of 10, with the same accuracy and an input impedance of $10k\Omega$. The latter has a minimum CMRR of 94dB and maximum V_{os} of $100\mu V$ and is stable into $1000pF$. Burr-Brown also offers a version with high-input common-mode voltage range ($\pm 200V$), as described later.

□ High-voltage differential amplifier

Figure 7.27 shows a clever method for increasing the common-mode input voltage range of the differencing amplifier circuit beyond the supply voltages without a corresponding reduction in differential gain. U_2 looks at the common-mode input signal at U_1 's input and removes it via R_5 and R_6 . Since there is no common-mode

signal left at either U_1 or U_2 , the CMRR of the op-amps is unimportant. The ultimate CMRR of this circuit is thus set by the matching of resistor ratios $R_1/R_5 = R_3/R_6$, with no great demands made on the accuracy of R_2 and R_4 . The circuit shown has a common-mode input range of ± 200 volts, a CMRR of 80dB, and a differential gain of 1.0.

Burr-Brown's unity-gain INA117 uses a different trick to achieve large common-mode voltage range, namely a 200:1 resistive voltage divider to bring the ± 200 volt signal within the op-amp's common-mode range of ± 10 volt (Fig. 7.28). This scheme is simpler than Figure 7.27, but suffers from degraded offset and noise specs: V_{os} is $1000\mu V$ (versus $250\mu V$ for the INA105), and output noise voltage is $25\mu V$ pp (0.01–10Hz) versus $2.4\mu V$ for the INA105.

□ Raising input impedance

The differencing circuit with carefully trimmed resistor values would seem to give the performance you want, until you look at the restrictions it puts on allowable source resistances. To get a gain accuracy of 0.1% with the circuit of Figure 7.26, you have to keep the source impedance below

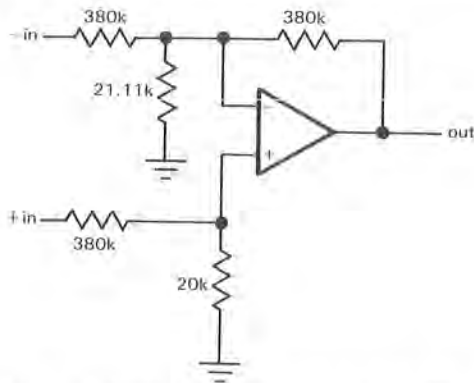


Figure 7.28. INA117 differential amplifier with ± 200 volts common-mode input range.

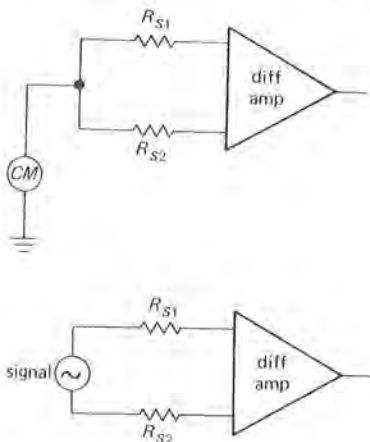


Figure 7.29

0.25 ohm! Furthermore, the source impedance seen at the two terminals has to be matched to 0.0025 ohm in order to attain a CMRR of 100dB. This last result follows from a look at the equivalent circuit (Fig. 7.29). The triangles represent the whole differential amplifier circuit or, in general, any differential or instrumentation amplifier, and R_{S1} and R_{S2} represent the Thévenin source resistances in each leg. For common-mode signals, the overall amplifier circuit includes the two source impedances in series with the input resistors R_1 and R_3 , and so the CMRR now depends on the matching of $R_{S1} + R_1$ with $R_{S2} + R_3$. Obviously the demands this circuit makes on the source impedances as calculated earlier are unreasonable.

Some improvement can be had by increasing the resistor values, using the trick of a T network for the feedback resistors, as in Figure 7.30. This is the differential amplifier version of the T network discussed in Sections 7.06 and 4.19. With the values shown, you get a differential voltage gain of 1000 (60dB). For a gain accuracy of 0.1%, the source impedance must be less than 25 ohms and must be matched to 0.25 ohm for 100dB CMRR. This is still an unacceptable demand on the source

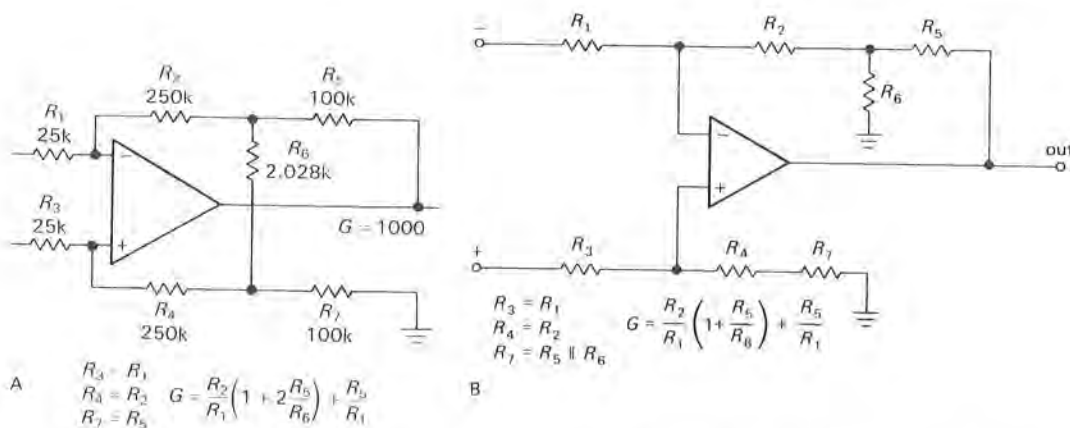


Figure 7.30. Differential amplifiers with T networks allow higher input impedances with smaller feedback resistors.

in most applications. A strain gauge, for instance, typically has a source impedance of about 350 ohms.

The general solution to this problem involves followers, or noninverting amplifiers, to attain high input impedance. The simplest method would be to add followers to the conventional differential amplifier (Fig. 7.31). With the enormous input impedances you get, there is no longer any problem with any reasonable source impedance, at least at dc. At higher frequencies it again becomes important to have matched source impedances relative to the common-mode signal, because the input capacitance of the circuit forms a voltage divider in combination with the source resistance. By "high frequencies" we often mean 60Hz, since common-mode ac power-line pickup is a common nuisance; at that frequency the effect of a few picofarads of input capacitance isn't serious.

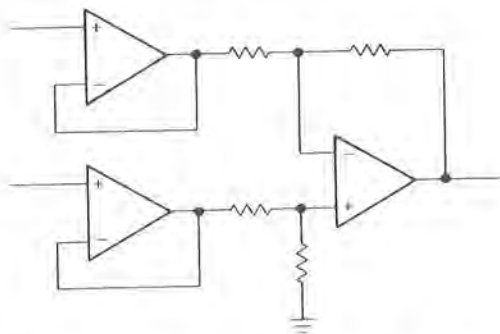


Figure 7.31. Differential amplifier with high Z_{in} .

7.10 Standard three-op-amp instrumentation amplifier

One disadvantage of the previous follower circuit (Fig. 7.31) is that it requires high CMRR both in the followers and in the final op-amp. Since the input buffers operate at unity gain, all the common-

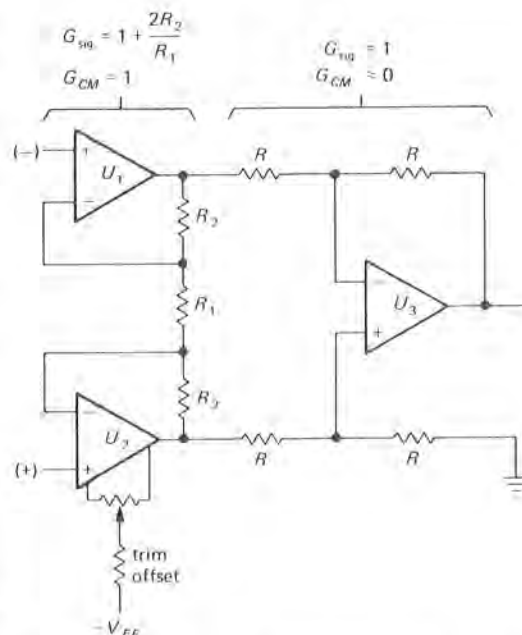


Figure 7.32. Classic instrumentation amplifier.

mode rejection must come in the output amplifier, requiring precise resistor matching, as we discussed. The circuit in Figure 7.32 is a significant improvement in this respect. It constitutes the standard instrumentation amplifier configuration. The input stage is a clever configuration of two op-amps that provides high differential gain and unity common-mode gain without any close resistor matching. Its differential output represents a signal with substantial reduction in the comparative common-mode signal, and it is used to drive a conventional differential amplifier circuit. The latter is often arranged for unity gain and is used to generate a single-ended output and polish off any remaining common-mode signal. As a result, the output op-amp, U_3 , needn't have exceptional CMRR itself, and resistor matching in U_3 's circuit is not terribly critical. Offset trimming for the whole circuit can be done at one of the input op-amps, as shown. The input op-amps must still have high CMRR, and they should be chosen carefully.

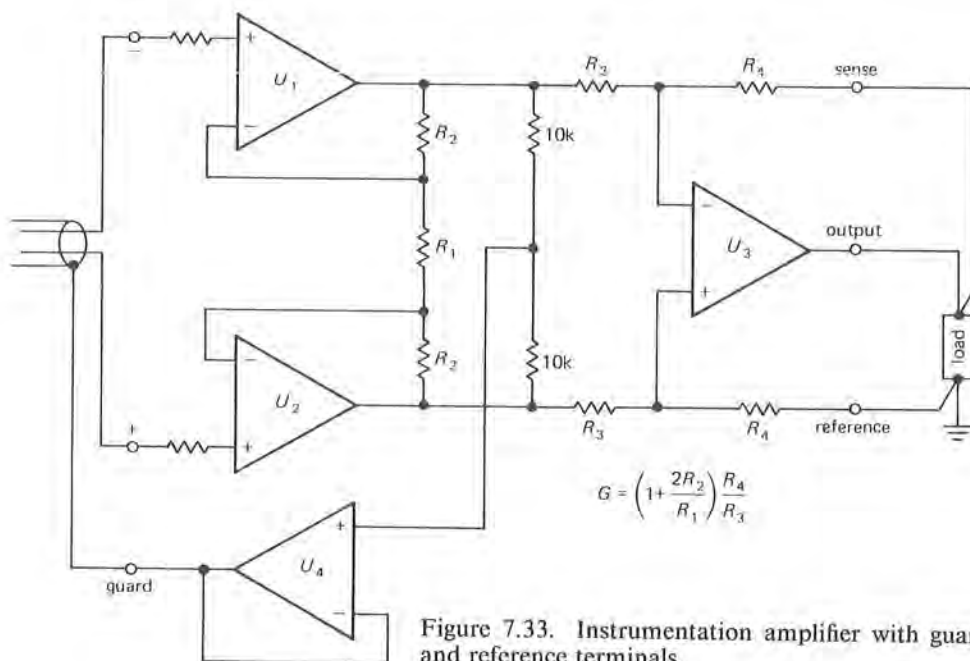


Figure 7.33. Instrumentation amplifier with guard, sense, and reference terminals.

Complete instrumentation amplifier ICs containing this standard configuration are available from several manufacturers. All components except R_1 are internal, with gain set by the single external resistor R_1 . Typical examples are the micropower INA102, high-speed INA110, and the high-accuracy AD624. All of these amplifiers offer a gain range of 1 to 1000, CMRR in the neighborhood of 100dB, and input impedances greater than 100M. The micropower hybrid LH0036 can run from supply voltages as low as ± 1 volt. The AD624 offers gain linearity of 0.001%, initial offset voltage of $25\mu\text{V}$, and offset drift of $0.25\mu\text{V}/^\circ\text{C}$, with provision for external trimming of offset voltage. Some instrumentation amplifiers (e.g., the high-accuracy INA104) have provision for CMRR trimming. Don't confuse these with the 725 "instrumentation operational amplifier," which is nothing more than a good op-amp intended as a building block for instrumentation amplifiers. Figure 7.33

shows the complete instrumentation amplifier circuit that is usually used.

A few comments about these instrumentation amplifier circuits (Fig. 7.33): (a) The buffered common-mode signal at U_4 's output can be used as a "guard" voltage to reduce the effects of cable capacitance and leakage. When used this way, the guard output will be tied to the shield of the input cables. If the gain-setting resistor (R_1) is not immediately adjacent to the amplifier (e.g., if it is a panel adjustment, a configuration that should usually be avoided), its connections should be shielded and guarded also. (b) The SENSE and REF terminals allow sensing of output voltage *at the load* so that feedback can operate to eliminate losses in the wiring or external circuit. In addition, the REF terminal also allows you to offset the output signal by a dc level (or by another signal); however, the impedance from the ref terminal to ground must be kept small, or the CMRR will be degraded. (c) With any of these

instrumentation amplifiers there must be a bias path for input current; for example, you can't just connect a thermocouple across the input. Figure 7.34 shows the simple application of an IC instrumentation amplifier with guard, sense, and reference terminals.

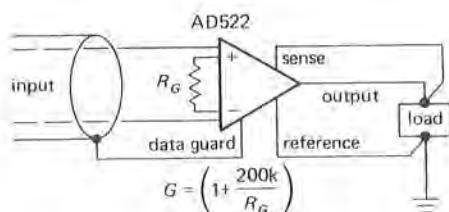


Figure 7.34. IC instrumentation amplifier.

□ Bootstrapped power supply

The CMRR of the input op-amps may be the limiting factor in the ultimate common-mode rejection of this circuit. If CMRRs greater than about 120dB are needed, the trick shown in Figure 7.35 can be used. U_4 buffers the common-mode signal level, driving the common terminal of a small floating split supply for U_1 and U_2 . This bootstrapping scheme effectively eliminates the input common-mode signal from U_1 and U_2 , because they see no swing (due to common-mode signals) at their inputs relative to their power supplies. U_3 and U_4 are powered by the system power supply, as usual. This scheme

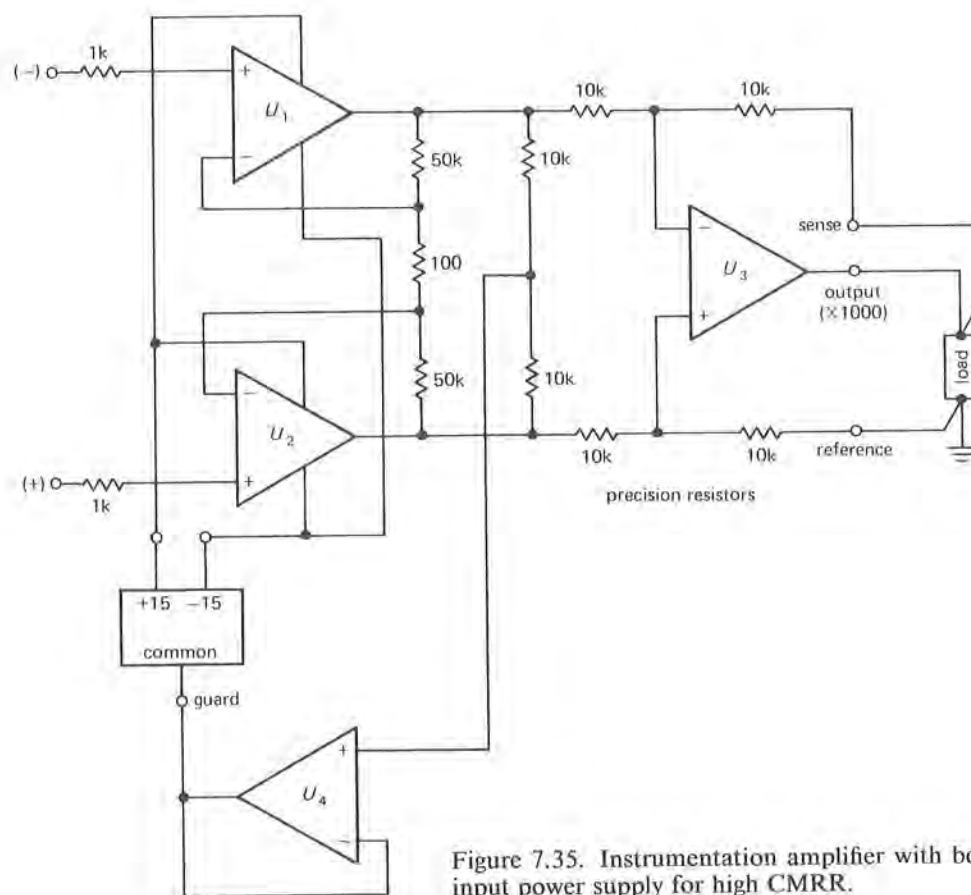


Figure 7.35. Instrumentation amplifier with bootstrapped input power supply for high CMRR.

can do wonders for the CMRR, at least at dc. At increasing frequencies you have the usual problems of presenting matched impedances to the input capacitances.

Two-op-amp configuration

Figure 7.36 shows another configuration that offers high input impedance with only two op-amps. Since it doesn't accomplish the common-mode rejection in two stages, as in the three-op-amp circuit, it requires precise resistor matching for good CMRR, in a manner similar to that of the standard differencing amplifier circuit.

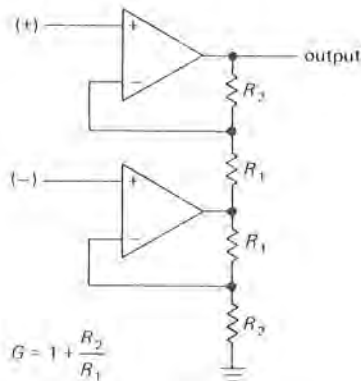


Figure 7.36. Instrumentation amplifier circuit with two op-amps.

Special IC instrumentation amplifiers

There are several interesting instrumentation amplifier configurations available as monolithic (and therefore inexpensive) ICs, some with extremely good performance. They use methods unrelated to the preceding circuits.

□ Current-feedback amplifier technique.

This technique, typified by the LM363, AD521, and JFET AMP-05, achieves high CMRR without the need for matched external resistors. In fact, the gain is set by the ratio of a pair of external

resistors. Figure 7.37 shows a block diagram of the AMP-01. The circuit employs two differential transconductance amplifier pairs, with a single external resistor setting the gain in each case. One pair is driven by the input signal, and the other is driven by the output signal, relative to the ref terminal. The AMP-05 uses FETs to keep input currents low, whereas the AMP-01 uses bipolar technology to achieve low offset voltage and drift (Table 7.5).

Computer-aided design methods can be extremely useful in precision circuit design; see Section 13.24.

AMPLIFIER NOISE

In almost every area of measurement the ultimate limit of detectability of weak signals is set by noise – unwanted signals that obscure the desired signal. Even if the quantity being measured is not weak, the presence of noise degrades the accuracy of the measurement. Some forms of noise are unavoidable (e.g., real fluctuations in the quantity being measured), and they can be overcome only with the techniques of *signal averaging* and *bandwidth narrowing*, which we will discuss in Chapter 15. Other forms of noise (e.g., radiofrequency interference and “ground loops”) can be reduced or eliminated by a variety of tricks, including filtering and careful attention to wiring configuration and parts location. Finally, there is noise that arises in the amplification process itself, and it can be reduced through the techniques of low-noise amplifier design. Although the techniques of signal averaging can often be used to rescue a signal buried in noise, it always pays to begin with a system that is free of preventable interference and that possesses the lowest amplifier noise practicable.

We will begin by talking about the origins and characteristics of the different

TABLE 7.5. INSTRUMENTATION AMPLIFIERS

Type	Noise																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
	Total supply			Maximum input errors ^b			Voltage																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
	Voltage		Curr minmax (mA)	Offset voltage		Current (nA)	0.1-10Hz		10Hz-10kHz																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
	(V)	(V)		RTI ^a (mV)	RTO ^a (μ V/°C)		RTI ^a (μ V, pp)	RTO ^a (μ V, rms)	RTI ^a (μ V, rms)	RTO ^a (pA, rms)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
CMRR (@dc, min)	G=1 (dB)	G=1k (dB)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)	G=1k (kHz)	G=1 (kHz)

(a) RTI: referred to the input; RTO: referred to the output. Noise and errors can be separated into components generated at both the input and output. The total input-referred noise (or error) is thus given by $RTI + RTO/G$. (b) diff¹ input impedance > 1G Ω except LH0038 (5M Ω), AMP-05 (1T Ω), and INA110B (5T Ω). (c) gain range 10-2000. (d) to 0.01%.

(e) CAZ type (see section 7.10); 7605 is uncomp. (f) G = 500. (g) G = 10. (h) 0.01Hz to 10Hz. (i) typical.

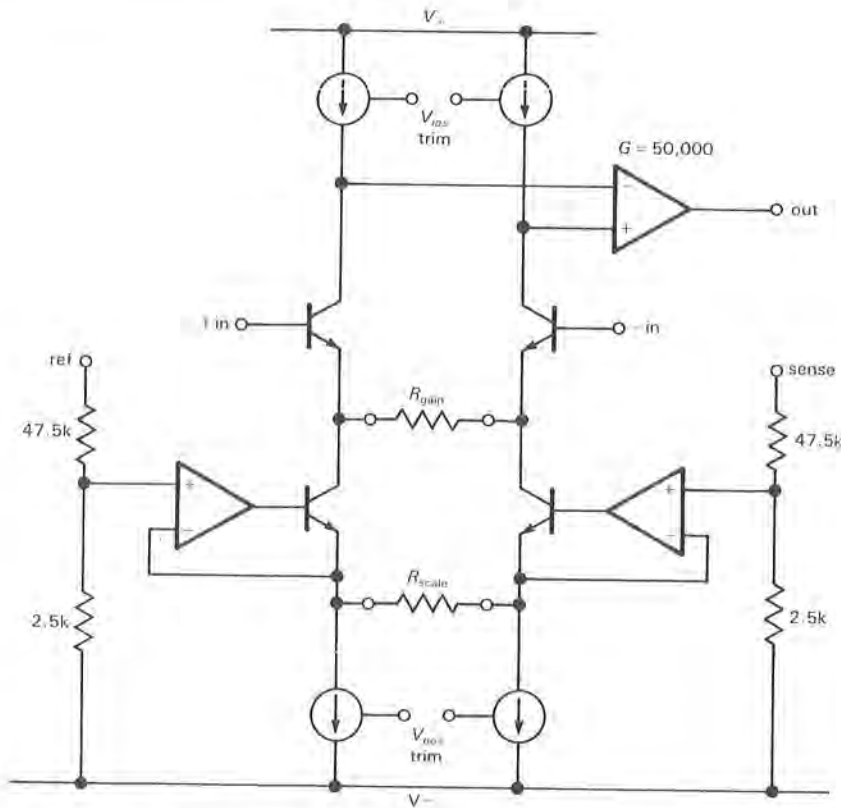


Figure 7.37. Block diagram of the AMP-01 instrumentation amplifier IC.

kinds of noise that afflict electronic circuits. Then we will launch into a discussion of transistor and FET noise, including methods for low-noise design with a given signal source, and will present some design examples. After a short discussion of noise in differential and feedback amplifiers, we will conclude with a section on proper grounding and shielding and the elimination of interference and pickup. See also Section 13.24 (Analog modeling tools).

7.11 Origins and kinds of noise

Since the term *noise* can be applied to anything that obscures a desired signal, noise can itself be another signal (“interference”); most often, however, we use

the term to describe “random” noise of a physical (often thermal) origin. Noise can be characterized by its frequency spectrum, its amplitude distribution, and the physical mechanism responsible for its generation. Let’s next look at the chief offenders.

Johnson noise

Any old resistor just sitting on the table generates a noise voltage across its terminals known as Johnson noise. It has a flat frequency spectrum, meaning that there is the same noise power in each hertz of frequency (up to some limit, of course). Noise with a flat spectrum is also called “white noise.” The actual open-circuit noise voltage generated by a resistance R at temperature T is given by

$$V_{\text{noise}}(\text{rms}) = V_n R = (4kTRB)^{\frac{1}{2}}$$

where k is Boltzmann's constant, T is the absolute temperature in degrees Kelvin ($^{\circ}\text{K} = ^{\circ}\text{C} + 273.16$), and B is the bandwidth in hertz. Thus, $V_{\text{noise}}(\text{rms})$ is what you would measure at the output if you drove a perfect noiseless bandpass filter (of bandwidth B) with the voltage generated by a resistor at temperature T . At room temperature ($68^{\circ}\text{F} = 20^{\circ}\text{C} = 293^{\circ}\text{K}$),

$$\begin{aligned} 4kT &= 1.62 \times 10^{-20} \text{V}^2/\text{Hz} - \Omega \\ (4kTR)^{\frac{1}{2}} &= 1.27 \times 10^{-10} R^{\frac{1}{2}} \quad \text{V/Hz}^{\frac{1}{2}} \\ &= 1.27 \times 10^{-4} R^{\frac{1}{2}} \quad \mu\text{V/Hz}^{\frac{1}{2}} \end{aligned}$$

For example, a 10k resistor at room temperature has an open-circuit rms voltage of $1.3\mu\text{V}$, measured with a bandwidth of 10kHz (e.g., by placing it across the input of a high-fidelity amplifier and measuring the output with a voltmeter). The source resistance of this noise voltage is just R . Figure 7.38 plots the simple relationship between Johnson-noise voltage density (rms voltage per square root bandwidth) and source resistance.

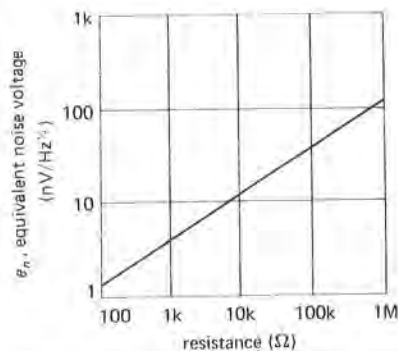


Figure 7.38. Thermal noise voltage versus resistance.

The amplitude of the Johnson-noise voltage at any instant is, in general, unpredictable, but it obeys a Gaussian amplitude distribution (Fig. 7.39),

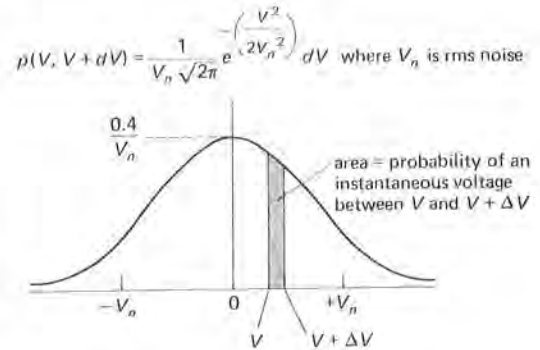


Figure 7.39

where $p(V)dV$ is the probability that the instantaneous voltage lies between V and $V + dV$, and V_n is the rms noise voltage, given earlier.

The significance of Johnson noise is that it sets a lower limit on the noise voltage in any detector, signal source, or amplifier having resistance. The resistive part of a source impedance generates Johnson noise, as do the bias and load resistors of an amplifier. You will see how it all works out shortly.

It is interesting to note that the physical analog of resistance (any mechanism of energy loss in a physical system, e.g., viscous friction acting on small particles in a liquid) has associated with it fluctuations in the associated physical quantity (in this case, the particles' velocity, manifest as the chaotic Brownian motion). Johnson noise is just a special case of this fluctuation-dissipation phenomenon.

Johnson noise should not be confused with the additional noise voltage created by the effect of resistance fluctuations when an externally applied current flows through a resistor. This "excess noise" has a $1/f$ spectrum (approximately) and is heavily dependent on the actual construction of the resistor. We will talk about it later.

Shot noise

An electric current is the flow of discrete electric charges, not a smooth fluidlike

flow. The finiteness of the charge quantum results in statistical fluctuations of the current. If the charges act independent of each other, the fluctuating current is given by

$$I_{\text{noise}}(\text{rms}) = I_{nR} = (2qI_{dc}B)^{\frac{1}{2}}$$

where q is the electron charge (1.60×10^{-19} coulomb) and B is the measurement bandwidth. For example, a “steady” current of 1 amp actually has an rms fluctuation of 57nA, measured in a 10kHz bandwidth; i.e., it fluctuates by about 0.000006%. The relative fluctuations are larger for smaller currents: A “steady” current of 1μA actually has an rms current noise fluctuation, measured over a 10kHz bandwidth, of 0.006%, i.e., -85dB. At 1pA dc, the rms current fluctuation (same bandwidth) is 56fA, i.e., a 5.6% variation! Shot noise is “rain on a tin roof.” This noise, like resistor Johnson noise, is Gaussian and white.

The shot-noise formula given earlier assumes that the charge carriers making up the current act independently. That is indeed the case for charges crossing a barrier, as for example the current in a junction diode, where the charges move by diffusion; but it is not true for the important case of metallic conductors, where there are long-range correlations between charge carriers. Thus, the current in a simple resistive circuit has far less noise than is predicted by the shot-noise formula. Another important exception to the shot-noise formula is provided by our standard transistor current-source circuit (Fig. 2.21), in which negative feedback acts to quiet the shot noise.

EXERCISE 7.4

A resistor is used as the collector load in a low-noise amplifier; the collector current I_C is accompanied by shot noise. Show that the output noise voltage is dominated by shot noise (rather than Johnson noise in the resistor) as long as the quiescent voltage drop

across the load resistor is greater than $2kT/q$ (50mV, at room temperature).

1/f noise (flicker noise)

Shot noise and Johnson noise are irreducible forms of noise generated according to physical principles. The most expensive and most carefully made resistor has exactly the same Johnson noise as the cheapest carbon resistor (of the same resistance). Real devices have, in addition, various sources of “excess noise.” Real resistors suffer from fluctuations in resistance, generating an additional noise voltage (which adds to the ever-present Johnson noise) proportional to the dc current flowing through them. This noise depends on many factors having to do with the construction of the particular resistor, including the resistive material and especially the end-cap connections. Here is a listing of typical excess noise for various resistor types, given as rms microvolts per volt applied across the resistor, measured over one decade of frequency:

Carbon-composition	0.10μV to 3.0μV
Carbon-film	0.05μV to 0.3μV
Metal-film	0.02μV to 0.2μV
Wire-wound	0.01μV to 0.2μV

This noise has approximately a $1/f$ spectrum (equal power per decade of frequency) and is sometimes called “pink noise.” Other noise-generating mechanisms often produce $1/f$ noise, examples being base current noise in transistors and cathode current noise in vacuum tubes. Curiously enough, $1/f$ noise is present in nature in unexpected places, e.g., the speed of ocean currents, the flow of sand in an hourglass, the flow of traffic on Japanese expressways, and the yearly flow of the Nile measured over the last 2000 years. If you plot the loudness of a piece of classical music versus time, you get a $1/f$ spectrum! No unifying principle has been found for all the $1/f$ noise that seems to be swirling around

us, although particular sources can often be identified in each instance.

Interference

As we mentioned earlier, an interfering signal or stray pickup constitutes a form of noise. Here the spectrum and amplitude characteristics depend on the interfering signal. For example, 60Hz pickup has a sharp spectrum and relatively constant amplitude, whereas car ignition noise, lightning, and other impulsive interferences are broad in spectrum and spiky in amplitude. Other sources of interference are radio and television stations (a particularly serious problem near large cities), nearby electrical equipment, motors and elevators, subways, switching regulators, and television sets. In a slightly different guise you have the same sort of problem generated by anything that puts a signal into the parameter you are measuring. For example, an optical interferometer is susceptible to vibration, and a sensitive radio-frequency measurement (e.g., NMR) can be affected by ambient radiofrequency signals. Many circuits, as well as detectors and even cables, are sensitive to vibration and sound; they are *microphonic*, in the terminology of the trade.

Many of these noise sources can be controlled by careful shielding and filtering, as we will discuss later in the chapter. At other times you are forced to take draconian measures, involving massive stone tables (for vibration isolation), constant-temperature rooms, anechoic chambers, and electrically shielded rooms.

7.12 Signal-to-noise ratio and noise figure

Before getting into the details of amplifier noise and low-noise design, we need to define a few terms that are often used to describe amplifier performance. These involve ratios of noise voltages, measured

at the same place in the circuit. It is conventional to refer noise voltages to the input of an amplifier (although the measurements are usually made at the output), i.e., to describe source noise and amplifier noise in terms of microvolts *at the input* that would generate the observed output noise. This makes sense when you want to think of the relative noise added by the amplifier to a given signal, independent of amplifier gain; it's also realistic, because most of the amplifier noise is usually contributed by the input stage. Unless we state otherwise, noise voltages are referred to the input.

Noise power density and bandwidth

In the preceding examples of Johnson noise and shot noise, the noise voltage you measure depends both on the measurement bandwidth B (i.e., how much noise you see depends on how fast you look) and on the variables (R and I) of the noise source itself. So it's convenient to talk about an rms noise-voltage "density" v_n :

$$V_n(\text{rms}) = v_n B^{\frac{1}{2}} = (4kTR)^{\frac{1}{2}} B^{\frac{1}{2}}$$

where V_n is the rms noise voltage you would measure in a bandwidth B . White-noise sources have a v_n that doesn't depend on frequency, whereas pink noise, for instance, has a v_n that drops off at 3dB/octave. You'll often see v_n^2 , too, the mean squared noise density. Since v_n always refers to rms, and v_n^2 always refers to mean square, you can just square v_n to get v_n^2 ! Sounds simple (and it is), but we want to make sure you don't get confused.

Note that B and the square root of B keep popping up. Thus, for example, for Johnson noise from a resistor R

$$v_{nR}(\text{rms}) = (4kTR)^{\frac{1}{2}} \quad \text{V/Hz}^{\frac{1}{2}}$$

$$v_{nR}^2 = 4kTR \quad \text{V}^2/\text{Hz}$$

$$V_n(\text{rms}) = v_{nR} B^{\frac{1}{2}} = (4kTRB)^{\frac{1}{2}} \quad \text{V}$$

$$V_n^2 = v_{nR}^2 B = 4kTRB \quad \text{V}^2$$

On data sheets you may see graphs of v_n or v_n^2 , with units like “nanovolts per root Hz” or “volts squared per Hz.” The quantities e_n and i_n that will soon appear work just the same way.

When you add two signals that are uncorrelated (two noise signals, or noise plus a real signal), the *squared* amplitudes add:

$$v = (v_s^2 + v_n^2)^{\frac{1}{2}}$$

where v is the rms signal obtained by adding together a signal of rms amplitude v_s and a noise signal of rms amplitude v_n . The rms amplitudes *don't* add.

Signal-to-noise ratio

Signal-to-noise ratio (SNR) is simply defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{V_s^2}{V_n^2} \right) \quad \text{dB}$$

where the voltages are rms values, and some bandwidth and center frequency are specified; i.e., it is the ratio, in decibels, of the rms voltage of the desired signal to the rms voltage of the noise that is also present. The “signal” itself may be sinusoidal or a modulated information-carrying waveform or even a noiselike signal itself. It is particularly important to specify the bandwidth if the signal has some sort of narrowband spectrum, since the SNR will drop as the bandwidth is increased beyond that of the signal: The amplifier keeps adding noise power, while the signal power remains constant.

Noise figure

Any real signal source or measuring device generates noise because of Johnson noise in its source resistance (the real part of its complex source impedance). There may be additional noise, of course, from other causes. The *noise figure* (NF) of an amplifier is simply the ratio, in decibels,

of the output of the real amplifier to the output of a “perfect” (noiseless) amplifier of the same gain, with a resistor of value R_s connected across the amplifier’s input terminals in each case. That is, the Johnson noise of R_s is the “input signal.”

$$\begin{aligned} \text{NF} &= 10 \log_{10} \left(\frac{4kTR_s + v_n^2}{4kTR_s} \right) \\ &= 10 \log_{10} \left(1 + \frac{v_n^2}{4kTR_s} \right) \quad \text{dB} \end{aligned}$$

where v_n^2 is the mean squared noise voltage per hertz contributed by the amplifier, with a noiseless (cold) resistor of value R_s connected across its input. This latter restriction is important, as you will see shortly, because the noise voltage contributed by an amplifier depends very much on the source impedance (Fig. 7.40).

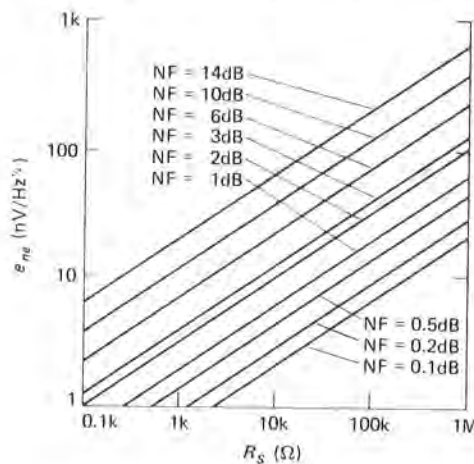


Figure 7.40. Effective noise voltage versus noise figure and source resistance. (National Semiconductor Corp.)

Noise figure is handy as a figure of merit for an amplifier when you have a signal source of a given source impedance and want to compare amplifiers (or transistors, for which NF is often specified). NF varies with frequency and source impedance, and it is often given as a set of contours of

constant NF versus frequency and R_s . It may also be given as a set of graphs of NF versus frequency, one curve for each collector current, or a similar set of graphs of NF versus R_s , one for each collector current. Note: The foregoing expressions for NF assume that the amplifier's input impedance is much larger than the source impedance, i.e., $Z_{in} \gg R_s$. However, in the special case of radiofrequency amplifiers, you usually have $R_s = Z_{in} = 50$ ohms, with NF defined accordingly. For this special case of matched impedances, simply remove the factors "4" from the foregoing equations.

Big fallacy: Don't try to improve things by adding a resistor in series with a signal source to reach a region of minimum NF. All you're doing is making the source noisier to make the amplifier look better! Noise figure can be very deceptive for this reason. To add to the deception, the NF specification (e.g., NF = 2dB) for a transistor or FET will always be for the optimum combination of R_s and I_C . It doesn't tell you much about actual performance, except that the manufacturer thinks the noise figure is worth bragging about.

In general, when evaluating the performance of some amplifier, you're probably least likely to get confused if you stick with SNR calculated for that source voltage and impedance. Here's how to convert from NF to SNR:

$$\text{SNR} = 10 \log_{10} \left(\frac{v_s^2}{4kTR_s} \right) - \text{NF(dB)} \quad (\text{at } R_s) \quad \text{dB}$$

where v_s is the rms signal amplitude, R_s is the source impedance, and NF is the noise figure of the amplifier for source impedance R_s .

□ Noise temperature

Rather than noise figure, you sometimes see noise temperature used to express

the noise performance of an amplifier. Both methods give the same information, namely the excess noise contribution of the amplifier when driven by a signal source of impedance R_s ; they are equivalent ways of expressing the same thing.

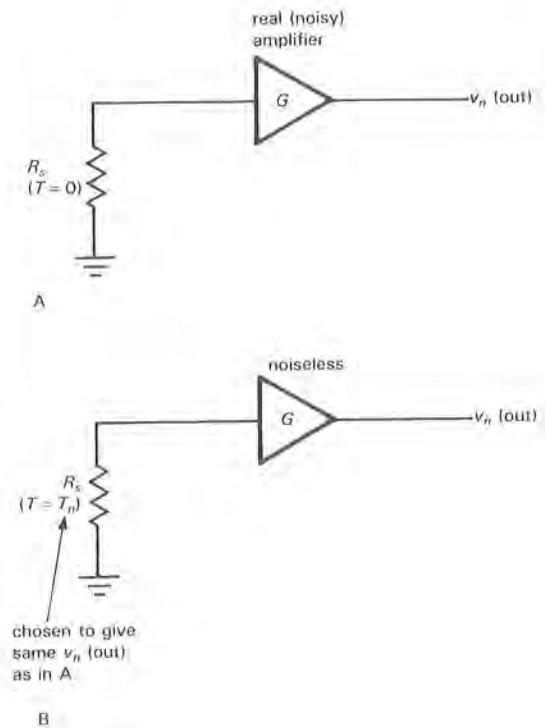


Figure 7.41

Look at Figure 7.41 to see how noise temperature works: We first imagine the actual (noisy) amplifier connected to a *noiseless* source of impedance R_s (Fig. 7.41A). If you have trouble imagining a noiseless source, think of a resistor of value R_s cooled to absolute zero. There will be some noise at the output, even though the source is noiseless, because the amplifier has noise. Now imagine constructing Figure 7.41B, where we magically make the amplifier noiseless, and bring the source R_s up to some temperature T_n such that the output noise voltage is the same as in Figure 7.41A. T_n is called the noise

temperature of the amplifier, for source impedance R_s .

As we remarked earlier, noise figure and noise temperature are simply different ways of conveying the same information. In fact, you can show that they are related by the following expressions:

$$T_n = T \left(10^{\text{NF(dB)}/10} - 1 \right)$$

$$\text{NF(dB)} = 10 \log_{10} \left(\frac{T_n}{T} + 1 \right)$$

where T is the ambient temperature, usually taken as 290°K.

Generally speaking, good low-noise amplifiers have noise temperatures far below room temperature (or, equivalently, they have noise figures far less than 3dB). Later in the chapter we will explain how you go about measuring the noise figure (or temperature) of an amplifier. First, however, we need to understand noise in transistors and the techniques of low-noise design. We hope the discussion that follows will clarify what is often a murky subject!

After reading the next two sections, we trust you won't ever be confused about noise figure again!

7.13 Transistor amplifier voltage and current noise

The noise generated by an amplifier is easily described by a simple noise model that is accurate enough for most purposes. In Figure 7.42, e_n represents a noise voltage source in series with the input, and i_n represents an input noise current. The transistor (or amplifier, in general) is assumed noiseless, and it simply amplifies the input noise voltage it sees. That is, the amplifier contributes a total noise voltage e_a , referred to the input, of

$$e_a(\text{rms}) = [e_n^2 + (R_s i_n)^2]^{\frac{1}{2}} \quad \text{V/Hz}^{\frac{1}{2}}$$

The two terms are simply the amplifier input noise voltage and the noise voltage

generated by the amplifier's input noise current passing through the source resistance. Since the two noise terms are usually uncorrelated, their squared amplitudes add to produce the effective noise voltage seen by the amplifier. For low source resistances the noise voltage e_n dominates, whereas for high source impedances the noise current i_n generally dominates.

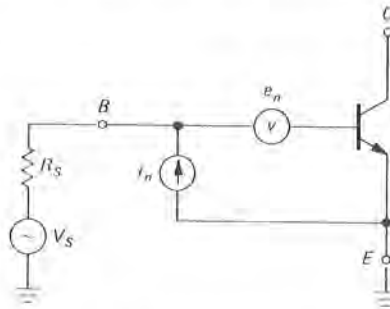


Figure 7.42. Noise model of a transistor.

Just to give an idea of what these look like, Figure 7.43 shows a graph of e_n and i_n versus I_C and f , for a 2N5087. We'll go into some detail now, describing these and showing how to design for minimum noise. It is worth noting that voltage noise and current noise for a transistor are in the range of nanovolts and picoamps per root hertz ($\text{Hz}^{\frac{1}{2}}$).

Voltage noise, e_n

The equivalent voltage noise looking in series with the base of a transistor arises from Johnson noise in the base spreading resistance, r_{bb} , and collector current shot noise generating a noise voltage across the intrinsic emitter resistance r_e . These two terms look like this:

$$e_n^2 = 4kTr_{bb} + 2qI_C r_e^2$$

$$= 4kTr_{bb} + \frac{2(kT)^2}{qI_C} \quad \text{V}^2/\text{Hz}$$

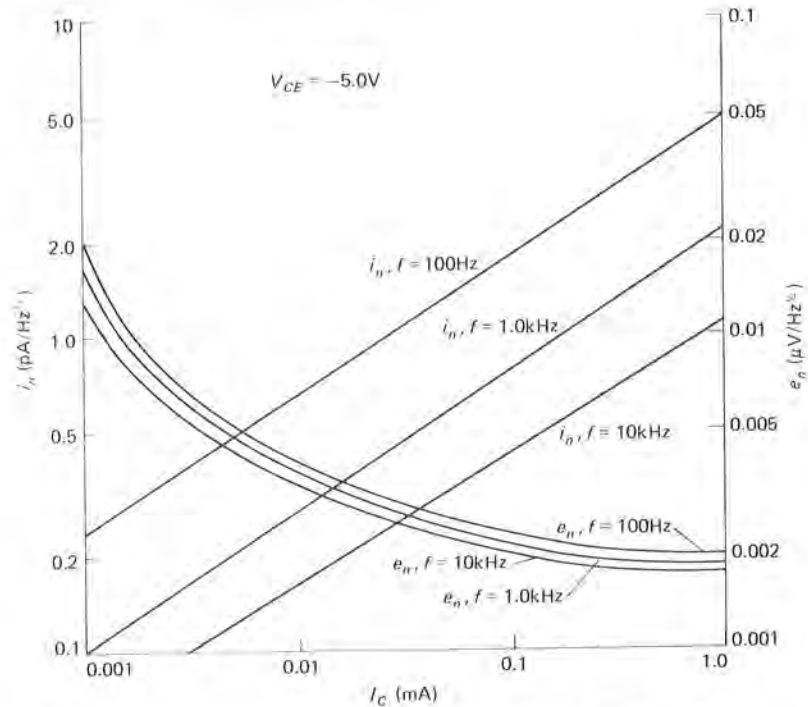


Figure 7.43. Equivalent rms input noise voltage (e_n) and noise current (i_n) versus collector current for a 2N5087 *npn* transistor. (Courtesy of Fairchild Camera and Instrument Corp.)

Both of these are Gaussian white noise. In addition, there is some flicker noise generated by base current flowing through r_{bb} . This last term is significant only at high base current, i.e., at high collector current. The result is that e_n is constant over a wide range of collector currents, rising at low currents (shot noise through an increasing r_e) and at sufficiently high currents (flicker noise from I_B through r_{bb}). This latter rise is present only at low frequencies, because of its $1/f$ character. As an example, at frequencies above 10kHz the 2N5087 has an e_n of $5\text{nV}/\text{Hz}^{1/2}$ at $I_C = 10\mu\text{A}$ and $2\text{nV}/\text{Hz}^{1/2}$ at $I_C = 100\mu\text{A}$. Figure 7.44 shows graphs of e_n versus frequency and current for the low-noise LM394 *npn* differential pair, and the low-noise 2SD786 from Toyo-Rohm. The latter uses special geometry to achieve an unusually low r_{bb} of 4 ohms, which is needed to realize the lowest values of e_n .

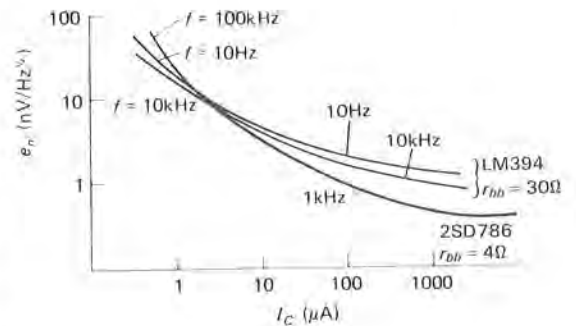


Figure 7.44. Input noise voltage (e_n) versus collector current for two low-noise bipolar transistors.

Current noise, i_n

Noise current is important, because it generates an additional noise voltage across the input signal source impedance. The main source of current noise is shot-noise fluctuation in the steady base current, added to the fluctuations caused by flicker

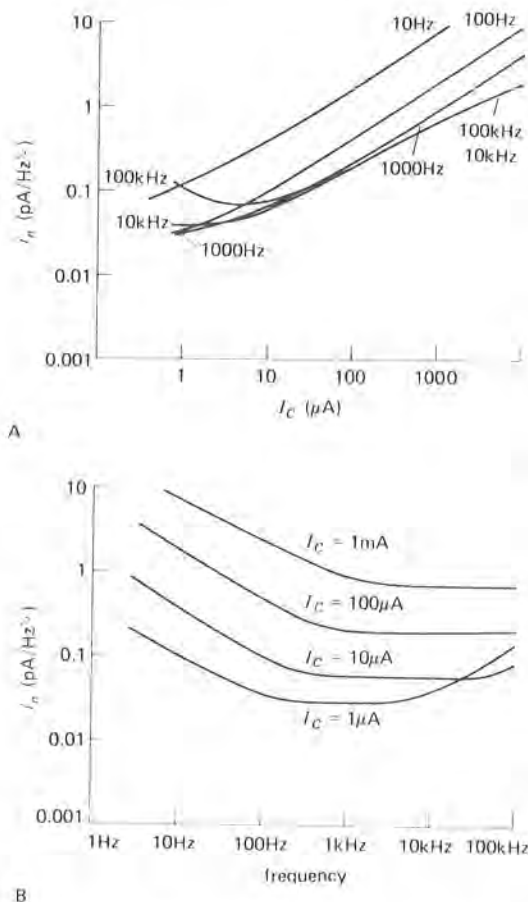


Figure 7.45. Input noise current for the LM394 bipolar transistor.
A. Noise current (i_n) versus collector current.
B. Noise current (i_n) versus frequency.

noise in r_{bb} . The shot-noise contribution is a noise current that increases proportional to the square root of I_B (or I_C) and is flat with frequency, whereas the flicker-noise component rises more rapidly with I_C and shows the usual $1/f$ frequency dependence. Taking the example of the 2N5087 again, above 10kHz i_n is about 0.1pA/Hz^{1/2} at $I_C = 10\mu\text{A}$ and 0.4pA/Hz^{1/2} at $I_C = 100\mu\text{A}$. The noise current increases, and the noise voltage drops, as I_C is increased. In the next section you will see how this dictates operating current in low-noise design.

Figure 7.45 shows graphs of i_n versus frequency and current, again for the low-noise LM394.

7.14 Low-noise design with transistors

The fact that e_n drops and i_n rises with increasing I_C provides a simple way to optimize transistor operating current to give lowest noise with a given source. Look at the model again (Fig. 7.46). The noiseless signal source v_s has added to it an irreducible noise voltage from the Johnson noise of its source resistance.

$$e_R^2(\text{source}) = 4kTR_s \quad \text{V}^2/\text{Hz}$$

The amplifier adds noise of its own, namely,

$$e_a^2(\text{amplifier}) = e_n^2 + (i_n R_s)^2 \quad \text{V}^2/\text{Hz}$$

Thus the amplifier's noise voltage is added to the input signal, and in addition, its noise current generates a noise voltage across the source impedance. These two are uncorrelated (except at very high frequencies), so you add their squares. The idea is to reduce the amplifier's total noise contribution as much as possible. That's easy, once you know R_s , because you just look at a graph of e_n and i_n versus I_C , in the region of the signal frequency, picking I_C to minimize $e_n^2 + (i_n R_s)^2$. Alternatively, if you are lucky and have a plot of noise-figure contours versus I_C and R_s , you can quickly locate the optimum value of I_C .

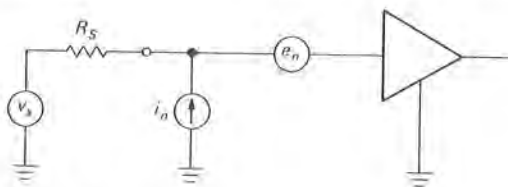


Figure 7.46. Amplifier noise model.

Noise figure example

As an example, suppose we have a small signal in the region of 1kHz with source

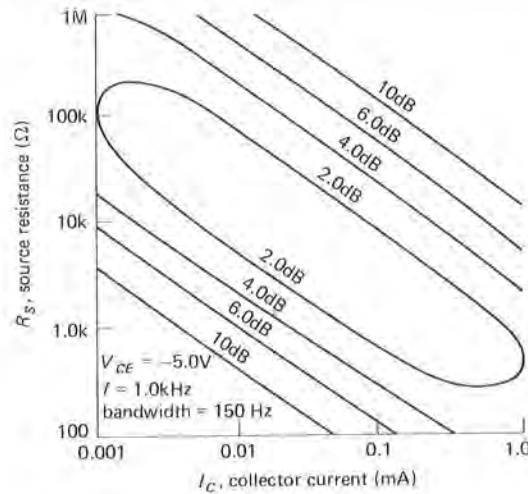


Figure 7.47. Contours of constant narrowband noise figure for the 2N5087 transistor. (Courtesy of Fairchild Camera and Instrument Corp.)

resistance of 10k, and we wish to make an amplifier with a 2N5087. From the e_n - i_n graph (Fig. 7.47) we see that the sum of voltage and current terms (with 10k source) is minimized for a collector current of about 10–20 μ A. Since the current noise is dropping faster than the voltage noise is rising as I_C is reduced, it might be a good idea to use slightly less collector current, especially if operation at a lower frequency is anticipated (i_n rises rapidly with decreasing frequency). We can estimate the noise figure using i_n and e_n at 1kHz:

$$\text{NF} = 10 \log_{10} \left(1 + \frac{e_n^2 + (i_n R_s)^2}{4kTR_s} \right) \text{ dB}$$

For $I_C = 10\mu\text{A}$, $e_n = 3.8\text{nV}/\text{Hz}^{1/2}$, $i_n = 0.29\text{pA}/\text{Hz}^{1/2}$, and $4kTR_s = 1.65 \times 10^{-16}\text{V}^2/\text{Hz}$ for the 10k source resistance. The calculated noise figure is therefore 0.6dB. This is consistent with the graph (Fig. 7.48) showing NF versus frequency, in which they have chosen $I_C = 20\mu\text{A}$ for $R_s = 10\text{k}$. This choice of collector current is also roughly what you would get from the graph in Figure 7.47 of noise-figure contours at 1kHz, although the

actual noise figure can be estimated only approximately from that plot as being less than 2dB.

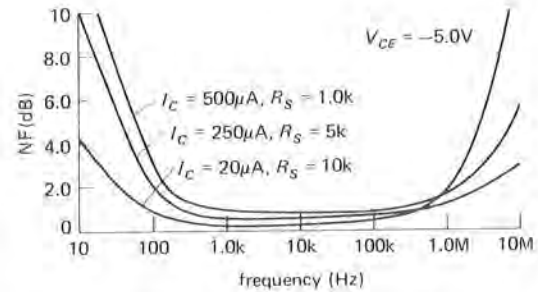


Figure 7.48. Noise figure (NF) versus frequency, for three choices of I_C and R_s , for the 2N5087. (Courtesy of Fairchild Camera and Instrument Corp.)

EXERCISE 7.5

Find the optimum I_C and corresponding noise figure for $R_s = 100\text{k}$ and $f = 1\text{kHz}$, using the graph in Figure 7.43 of e_n and i_n . Check your answer from the noise-figure contours (Fig. 7.47).

For the other amplifier configurations (follower, grounded base) the noise figure is essentially the same, for given R_s and I_C , since e_n and i_n are unchanged. Of course, a stage with unity voltage gain (a follower) may just pass the problem along to the next stage, since the signal level hasn't been increased to the point that low-noise design can be ignored in subsequent stages.

□ Charting amplifier noise with e_n and i_n

The noise calculations just presented, although straightforward, make the whole subject of amplifier design appear somewhat formidable. If you misplace a factor of Boltzmann's constant, you suddenly get an amplifier with 10,000dB noise figure! In this section we will present a simplified noise-estimation technique of great utility.

The method consists of first choosing some frequency of interest in order to get

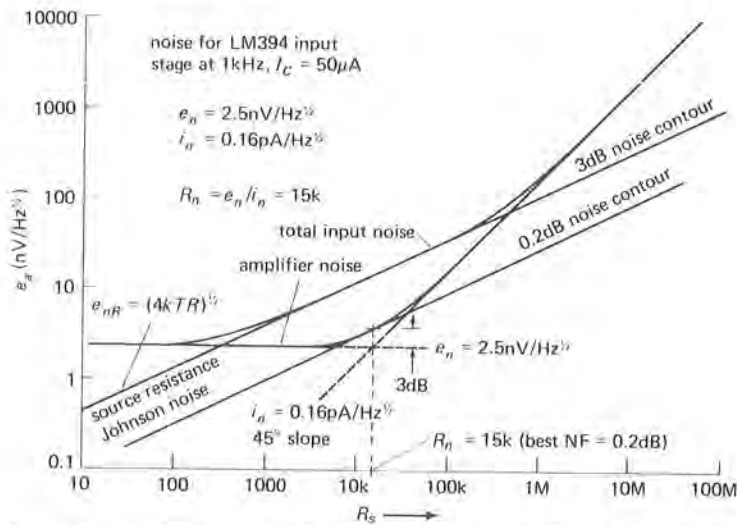


Figure 7.49. Total amplifier input voltage noise (e_a) plotted from the e_n and i_n parameters.

values for e_n and i_n versus I_C from the transistor data sheets. Then, for a given collector current, you can plot the total noise contributions from e_n and i_n as a graph of e_a versus source resistance R_s . Figure 7.49 shows what that looks like at 1kHz for a differential input stage using an LM394 matched superbeta transistor running 50 μ A of collector current. The e_n noise voltage is constant, and the $i_n R_s$ voltage increases proportional to R_s , i.e., with a 45° slope. The amplifier noise curve is drawn as shown, with care being taken to ensure that it passes through a point 3dB (voltage ratio of 1.4) above the crossing point of individual voltage and current noise contributions. Also plotted is the noise voltage of the source resistance, which also happens to be the 3dB NF contour. The other lines of constant noise figure are simply straight lines parallel to this line, as you will see in the examples that follow.

The best noise figure (0.2dB) at this collector current and frequency occurs for a source resistance of 15k, and the noise figure is easily seen to be less than 3dB for all source resistances between 300 ohms

and 500k, the points at which the 3dB NF contour intersects the amplifier noise curve.

The next step is to draw a few of these noise curves on the same graph, using different collector currents or frequencies, or maybe a selection of transistor types, in order to evaluate amplifier performance. Before we go on to do that, let's show how we can talk about this same amplifier using a different pair of noise parameters, the noise resistance R_n and the noise figure $NF(R_N)$, both of which pop right out of the graph.

□ Noise resistance

The lowest noise figure in this example occurs for a source resistance $R_s = 15k$, which equals the ratio of e_n to i_n . That defines the noise resistance

$$R_n = \frac{e_n}{i_n}$$

You can find the noise figure for a source of that resistance from our earlier expression for noise figure. It is

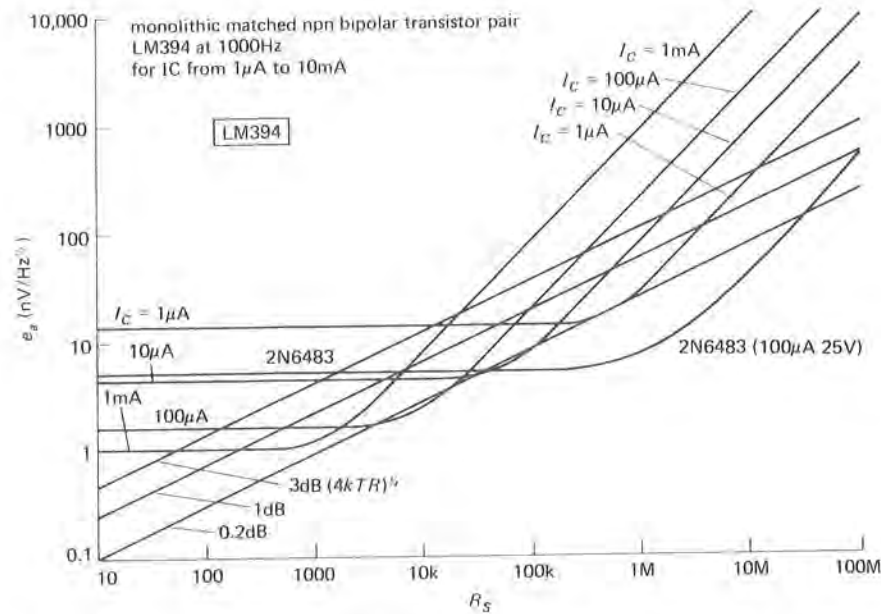


Figure 7.50. Total amplifier input voltage noise (e_n) for the LM394 bipolar transistor under various conditions, compared with the 2N6483 JFET.

$$\begin{aligned} \text{NF (at } R_n) &= \\ 10 \log_{10} \left(1 + 1.23 \times 10^{20} \frac{e_n^2}{R_n} \right) &\text{ dB} \\ \approx 0.2 &\text{ dB} \end{aligned}$$

Noise resistance isn't actually a real resistance in the transistor, or anything like that. It is a tool to help you quickly find the value of source resistance for minimum noise figure, ideally so that you can vary the collector current to shift R_n close to the value of source resistance you're actually using. R_n corresponds to the point where the e_n and i_n lines cross.

The noise figure for a source resistance equal to R_n then follows simply from the preceding equation.

□ Charting the bipolar/FET shootout

Let's have some fun with this technique. A perennial bone of contention among engineers is whether FETs or bipolar transistors are "better." We will dispose of

this issue with characteristic humility by matching two of the best contenders and letting them deliver their best punches. In the interest of fairness, we'll let National Semiconductor intramural teams compete, choosing two game fighters.

In the bipolar corner we have the magnificent LM394 superbeta monolithic matched pair, already warmed up, as described earlier. We'll run it at 1kHz, with collector currents from 1μA to 1mA (Fig. 7.50).

The FET entry is the 2N6483 monolithic n -channel JFET matched pair, known far and wide for its stunning low-noise performance, reputed to exceed that of bipolar transistors. According to its data sheet, it was trained only for 100μA and 400μA drain currents (Fig. 7.51).

And the winner? Well, it's a split decision. The FET won points on lowest minimum noise figure, $\text{NF}(R_n)$, reaching a phenomenal 0.05dB noise figure, and dipping well below 0.2dB from 100k to 100M source impedance. For high source

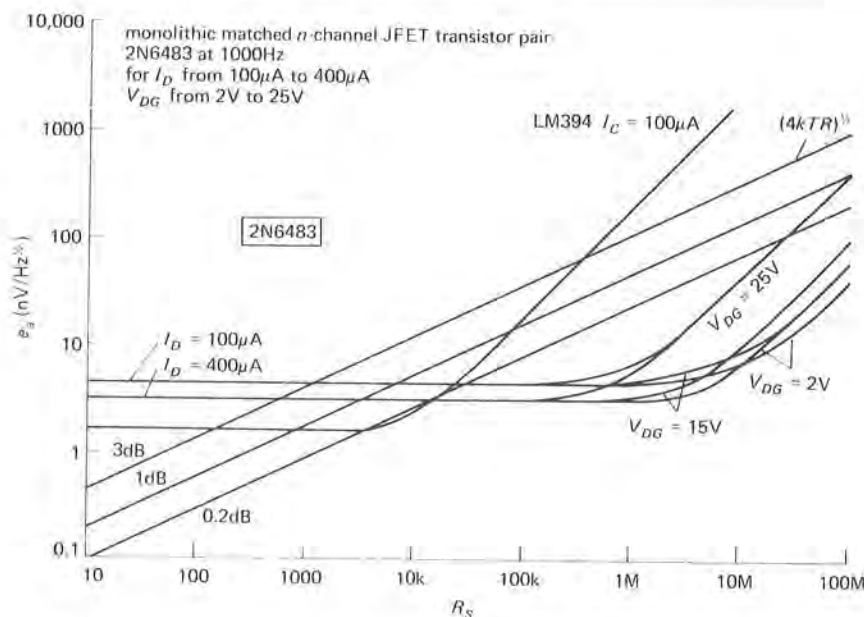


Figure 7.51. Total amplifier input voltage noise (e_a) for the 2N6483 JFET compared with the LM394 bipolar transistor.

impedances, FETs remain unbeaten. The bipolar transistor is best at low source impedances, particularly below 5k, and it can reach a 0.3dB noise figure at $R_s = 1k$, with suitable choice of collector current. By comparison, the FET cannot do better than 2dB with a 1k source resistance, owing to larger voltage noise e_n .

Just as in boxing, where the best fighters haven't yet had a chance to compete in a world championship, there are some younger contenders for the best low-noise transistor. For example, the 2SJ72 and 2SK147 complementary JFETs from Toshiba use a meshed-gate geometry to achieve a phenomenal e_n of $0.7nV/\sqrt{Hz}$ at $I_D = 10mA$ (equivalent to Johnson noise from a 30Ω resistor!). But these are JFETs, with their low input current (hence low i_n), and thus the noise resistance is about 10k. When used as an amplifier with a source impedance equal to their noise resistance (i.e., $R_s = 10k$), their performance is unbeatable – the noise temperature is just $2^\circ K$!

Before you go out and buy a bushel of these remarkable JFETs, consider the remarks of the critics, who claim they are muscle-bound – they have high input and feedback capacitance (85pF and 15pF, respectively), which limits their usefulness at high frequencies. Their relative, the 2SK117, is better in this regard, at the expense of higher e_n . These same critics argue that the Toyo-Rohm bipolar complementary pair, the 2SD786 and 2SB737, with e_n as low as $0.55nV/\sqrt{Hz}$, can offer even better performance at moderate source impedances and frequencies.

□ Low source impedances

Bipolar transistor amplifiers can provide very good noise performance over the range of source impedances from about 200 ohms to 1M; corresponding optimum collector currents are generally in the range of several milliamps down to a microamp. That is, collector currents used for the input stage of low-noise amplifiers generally

tend to be lower than in amplifier stages not optimized for low-noise performance.

For very low source impedances (say 50Ω), transistor voltage noise will always dominate, and noise figures will be poor. The best approach in such cases is to use a transformer to raise the signal level (and impedance), treating the signal on the secondary as before. High-quality signal transformers are available from companies such as James and Princeton Applied Research. As an example, the latter's model 116 FET preamp has voltage and current noise such that the lowest noise figure occurs for signals of source impedance around 1M . A signal around 1kHz with source impedance of 100 ohms would be a poor match for this amplifier, since the amplifier's voltage noise is much larger than the signal source's Johnson noise; the resultant noise figure for that signal connected directly to the amplifier would be 11dB . By using the optional internal step-up transformer, the signal level is raised (along with its source impedance), thus overriding amplifier noise voltage and giving a noise figure of about 1.0dB .

At radiofrequencies (e.g., beginning around 100kHz) it is extremely easy to make good transformers, both for tuned (narrowband) and broadband signals. At these frequencies it is possible to make broadband "transmission-line transformers" of very good performance. We will treat some of these methods in Chapter 13. It is at the very low frequencies (audio and below) that transformers become problematic.

Three comments: (a) The voltage rises proportional to the turns ratio of the transformer, whereas the impedance rises proportional to the square of the ratio. Thus a $2:1$ voltage step-up transformer has an output impedance four times the input impedance (this is mandated by conservation of energy). (b) Transformers aren't perfect. They have trouble at low frequencies (magnetic saturation) and at high frequencies

(winding inductance and capacitance), as well as losses from the magnetic properties of the core and from winding resistance. The latter is a source of Johnson noise, as well. Nevertheless, when dealing with a signal of very low source impedance, you may have no choice, and transformer coupling can be very beneficial, as the preceding example demonstrates. Exotic techniques such as cooled transformers, superconducting transformers, and SQUIDs (superconducting quantum interference devices) can provide good noise performance at low impedance and voltage levels. With SQUIDs you can measure voltages of 10^{-15} volt! (c) Again, a warning: Don't attempt to improve performance by adding a resistor in series with a low source impedance. If you do that, you're just another victim of the noise-figure fallacy.

□ High source impedances

If the source impedance is high, say greater than 100k or so, transistor current noise dominates, and the best device for low-noise amplification is a FET. Although their voltage noise is usually greater than that of bipolar transistors, the gate current (and its noise) can be exceedingly small, making them ideally suited for low-noise high-impedance amplifiers. Incidentally, it is sometimes useful to think of Johnson noise as a current noise $i_n = v_n/R_s$. This lets you compare source noise contributions with amplifier current noise (Fig. 7.52).

7.15 FET noise

We can use the same amplifier noise model for FETs, namely a series noise voltage source and a parallel noise current source. You can analyze the noise performance with exactly the same methods used for bipolar transistors. For example, see the graphs in the section on bipolar/FET shootout.

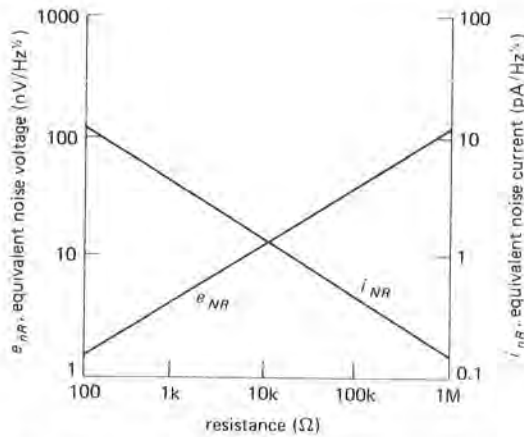


Figure 7.52. Thermal noise voltage density versus resistance at 25°C. The equivalent short-circuit current noise density is also shown.

Voltage noise of JFETs

For JFETs the voltage noise e_n is essentially the Johnson noise of the channel resistance, given approximately by

$$e_n^2 = 4kT \left(\frac{2}{3} \frac{1}{g_m} \right) \quad \text{V}^2/\text{Hz}$$

where the inverse transconductance takes the place of resistance in the Johnson-noise formula. Since the transconductance rises with increasing drain current (as $\sqrt{I_D}$), it is generally best to operate FETs at high drain current for lowest voltage noise. However, since the e_n is Johnson noise, which goes only as $1/\sqrt{g_m}$, and that in turn goes as $\sqrt{I_D}$, e_n is finally proportional to $I_D^{-1/4}$. With such a mild dependence of e_n on I_D it doesn't pay to run at a drain current so high that other properties of the amplifier are degraded. In particular, a FET running at high current gets hot, which (a) decreases g_m , (b) increases offset voltage drift and CMRR, and (c) raises gate leakage dramatically; the latter effect can actually *increase* voltage noise, since there is some contribution to e_n from flicker noise associated with the gate leakage current.

There is another way to increase g_m , and therefore decrease JFET voltage noise: By paralleling a pair of JFETs you get twice the g_m , but of course this is at twice the I_D . But now if you run the combination at the previous I_D , you still improve g_m by a factor of $\sqrt{2}$ over the single-JFET value, without increasing total drain current. In practice you can simply parallel a number of matched JFETs, or look for a large-geometry JFET like the 2SJ72 and 2SK147 mentioned earlier.

There is a price to pay, however. All the capacitances scale with the number of paralleled JFETs. As a result, high-frequency performance (including noise figure) is degraded. In practice you should stop paralleling additional transistors when the circuit's input capacitance roughly matches the source's capacitance. If you care about performance at high frequencies, choose JFETs with high g_m and low C_{rss} ; you might consider the ratio g_m/C_{rss} a high-frequency figure of merit. Note that circuit configurations can also play an important role; e.g., the cascode circuit can be used to eliminate the Miller effect (gain multiplication) on C_{rss} .

MOSFETs tend to have much higher voltage noise than JFETs, with $1/f$ noise predominating, since the $1/f$ knee is as high as 10kHz to 100kHz. For this reason you wouldn't normally choose a MOSFET for low-noise amplifiers below 1MHz.

Current noise of JFETs

At low frequencies the current noise i_n is extremely small, arising from the shot noise in the gate leakage current (Fig. 7.53):

$$i_n = (3.2 \times 10^{-19} I_G B)^{\frac{1}{2}} \quad \text{A(rms)}$$

In addition, there is a flicker-noise component in some FETs. The noise current rises with increasing temperature, as the gate leakage current rises. Watch out for the rapidly increasing gate leakage in *n*-channel JFETs that occurs for operation at high V_{DG} (see Section 3.09).

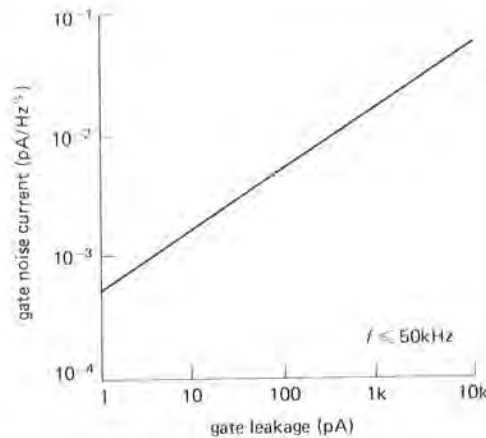


Figure 7.53. Input noise current versus gate leakage current for JFETs. (Courtesy of National Semiconductor Corp.)

At moderate to high frequencies there is an additional noise term, namely the real part of the input impedance seen looking into the gate. This comes from the effect of feedback capacitance (Miller effect) when there is a phase shift at the output due to load capacitance; i.e., the part of the output signal that is shifted 90° couples through the feedback capacitance C_{rss} to produce an effective resistance at the input, given by

$$R = \frac{1 + \omega C_L R_L}{\omega^2 g_m C_{rss} C_L R_L^2} \quad \text{ohms}$$

As an example, the 2N5266 *p*-channel JFET has a noise current of 0.005 pA/Hz^{1/2} and a noise voltage e_n of 12 nV/Hz^{1/2}, both at I_{DSS} and 10 kHz. The noise current begins climbing at about 50 kHz. These figures are roughly 100 times better in i_n and 5 times worse in e_n than the corresponding figures for the 2N5087 used earlier.

With FETs you can achieve good noise performance for input impedances in the range of 10 k to 100 M. The PAR model 116 preamp has a noise figure of 1 dB or better for source impedances from 5 k to 10 M in the frequency range from 1 kHz to 10 kHz. Its performance at moderate frequencies

corresponds to a noise voltage of 4 nV/Hz^{1/2} and a noise current of 0.013 pA/Hz^{1/2}.

7.16 Selecting low-noise transistors

As we mentioned earlier, bipolar transistors offer the best noise performance with low source impedances, owing to their lower input voltage noise. Voltage noise, e_n , is reduced by choosing a transistor with low base spreading resistance, r_{bb} , and operating at high collector current (as long as h_{FE} remains high). For higher source impedances the current noise can be minimized instead by operating at lower collector current.

At high values of source impedance, FETs are the best choices. Their voltage noise can be reduced by operating at higher drain currents, where the transconductance is highest. FETs intended for low-noise applications have high k values (see Section 3.04), which usually means high input capacitance. For example, the low-noise 2N6483 has $C_{iss} = 20$ pF, whereas the 2N5902 low-current FET has $C_{iss} = 2$ pF.

Figures 7.54 and 7.55 show comparisons of the noise characteristics of a number of popular and useful transistors.

□ 7.17 Noise in differential and feedback amplifiers

Low-noise amplifiers are often differential, to obtain the usual benefits of low drift and good common-mode rejection. When you calculate the noise performance of a differential amplifier, there are three points to keep in mind: (a) Be sure to use the individual collector currents, not the sum, to get e_n and i_n from data sheets. (b) The i_n seen at each input terminal is the same as for a single-ended amplifier configuration. (c) The e_n seen at one input, with the other input grounded, say, is 3 dB

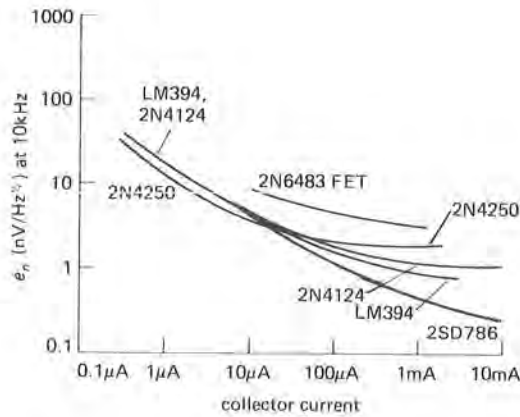
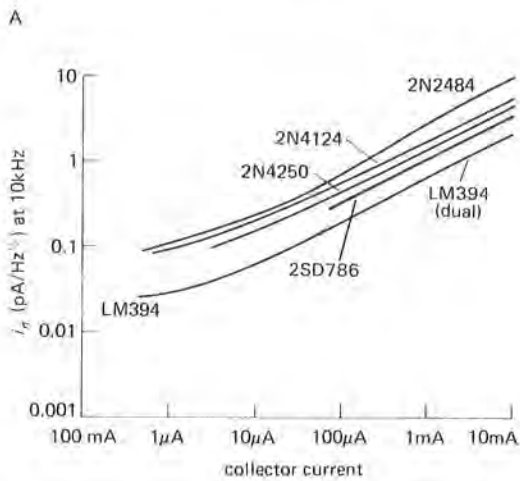
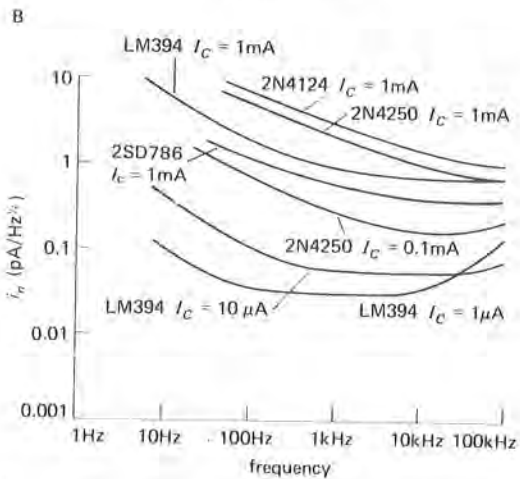
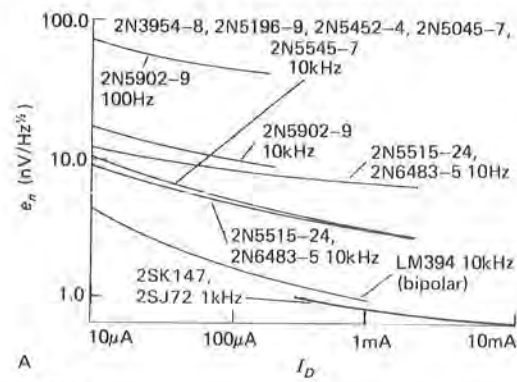
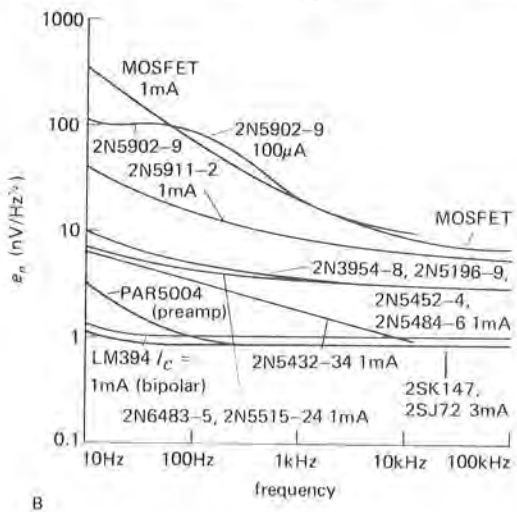
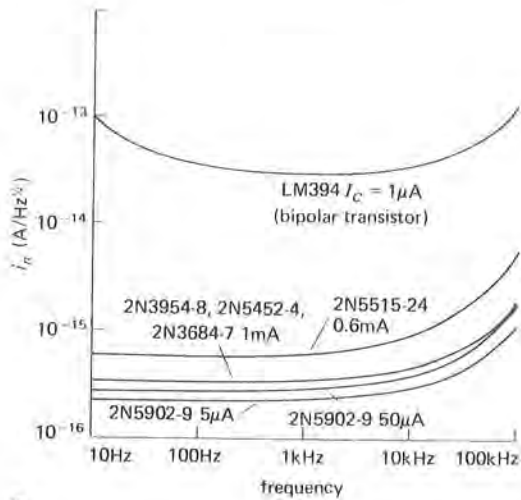


Figure 7.54. Input noise for some popular transistors.

A. Input noise voltage (e_n) versus collector current.

B. Input noise current (i_n) versus collector current.

C. Input noise current (i_n) versus frequency.

Figure 7.55. Input noise for some popular FETs.

A. Input noise voltage (e_n) versus drain current (I_D).

B. Input noise voltage (e_n) versus frequency.

C. Input noise current (i_n) versus frequency.

larger than the single-transistor case, i.e., it is multiplied by $\sqrt{2}$.

In amplifiers with feedback, you want to take the equivalent noise sources e_n and i_n out of the feedback loop, so you can use them as previously described when calculating noise performance with a given signal source. Let's call the noise terms brought out of the feedback loop e_A and i_A , for *amplifier* noise terms. Thus the amplifier's noise contribution to a signal with source resistance R_s is

$$e^2 = e_A^2 + (R_s I_A)^2 \quad \text{V}^2/\text{Hz}$$

Let's take the two feedback configurations separately.

□ Noninverting

For the noninverting amplifier (Fig. 7.56) the input noise sources become

$$\begin{aligned} i_A^2 &= i_n^2 \\ e_A^2 &= e_n^2 + 4kTR_{\parallel} + (i_n R_{\parallel})^2 \end{aligned}$$

where e_n is the "adjusted" noise voltage for the differential configuration, i.e., 3dB larger than for a single-transistor stage. The additional noise voltage terms arise from Johnson noise and input-stage noise current in the feedback resistors. Note that the effective noise voltage and current are now not completely uncorrelated, so calculations in which their squares are added can be in error by a maximum factor of 1.4.

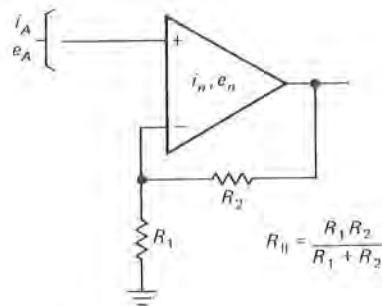


Figure 7.56

For a follower, R_2 is zero, and the effective noise sources are just those of the differential amplifier alone.

□ Inverting

For the inverting amplifier (Fig. 7.57) the input noise sources become

$$\begin{aligned} i_A^2 &= i_n^2 + 4kT \frac{1}{R_2} \\ e_A^2 &= e_n^2 + R_1^2 \left(i_n^2 + 4kT \frac{1}{R_2} \right) \\ &= e_n^2 + R_1^2 i_A^2 \end{aligned}$$

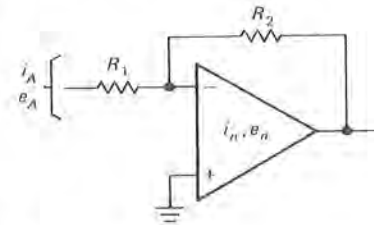


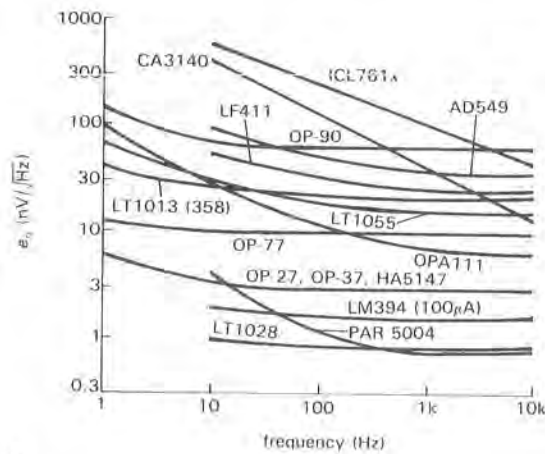
Figure 7.57

Op-amp selection curves

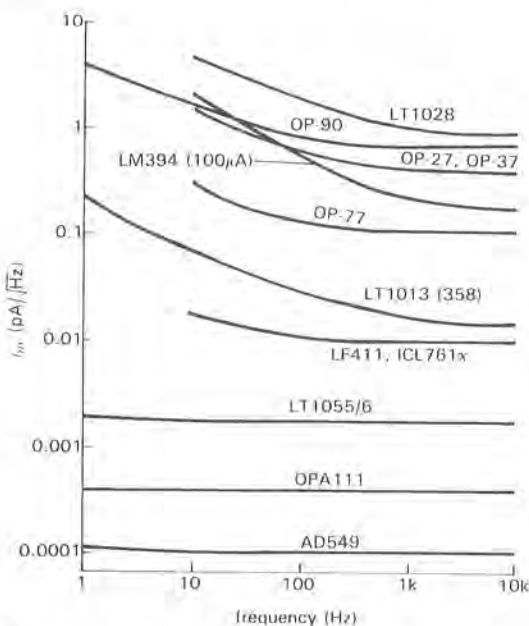
You now have all the tools necessary to analyze op-amp input circuits. Their noise is specified in terms of e_n and i_n , just as with transistors and FETs. You don't get to adjust anything, though; you only get to use them. The data sheets may need to be taken with a grain of salt. For example, "popcorn noise" is typified by jumps in offset at random times and duration. It is rarely mentioned in polite company. Figure 7.58 summarizes the noise performance of some popular op-amps.

Wideband noise

Op-amp circuits are generally dc-coupled and extend to some upper frequency limit f_{cutoff} . Therefore it is of interest to know the total noise voltage over this band, not merely the noise power density. Figure 7.59 presents some graphs showing the rms noise voltage in a band extending



A



B

Figure 7.58. Input noise for some popular op-amps.

A. Input noise voltage (e_n) versus frequency.
B. Input noise current (i_n) versus frequency.

from dc to the indicated frequency; they were calculated by integrating the noise power curves for the various op-amps.

Choosing a low-noise op-amp

It is simple to choose an op-amp to minimize noise in some frequency range, given

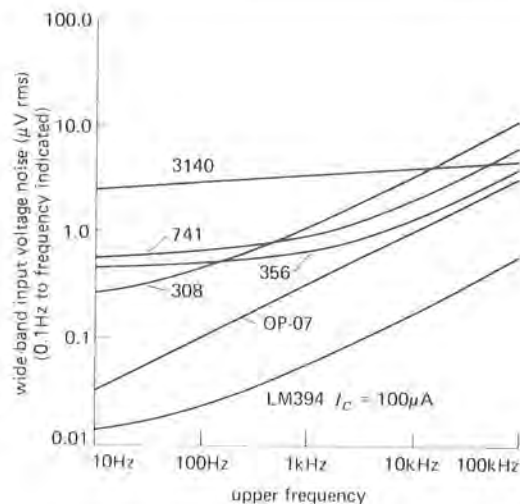


Figure 7.59. Wideband noise voltage for some popular op-amps.

the signal impedance seen from the op-amp, R_{sig} (which includes the effects of feedback components, as given in the foregoing expressions). Generally speaking, you want op-amps with low i_n for high signal impedances, and op-amps with low e_n for low signal impedances. Assuming the signal source is at room temperature, the total input-referred squared noise voltage density is just

$$e_A^2 = 4kTR_{sig} + e_n^2 + (i_n R_{sig})^2$$

where the first term is due to Johnson noise, and the last two terms are due to op-amp noise voltage and current. Obviously the Johnson noise sets a lower bound to the input-referred noise. In Figure 7.60 we've plotted the quantity e_A (at 10Hz) as a function of R_{sig} for the quietest op-amps we could find. For comparison we included our jellybean JFET LF411 and the micropower bipolar OP-90. The latter, although an excellent micropower op-amp, has high noise voltage (because the front end operates at low collector current, hence high r_e and therefore high Johnson noise) and also high noise current (because the bipolar input has substantial base current); it shows just how good

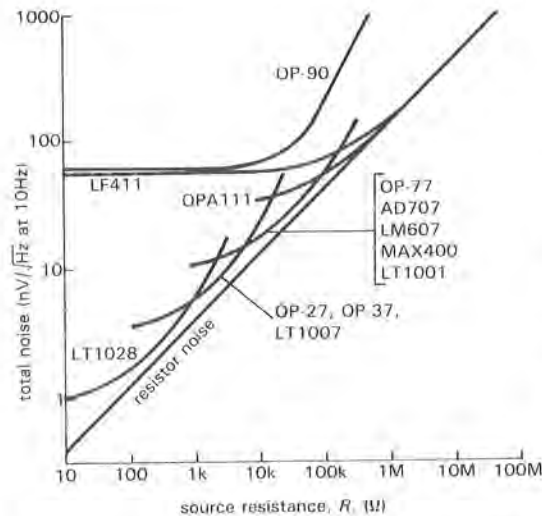


Figure 7.60. Total noise (source resistor plus amplifier, at 10Hz) for high-performance op-amps.

current of an amplifier, and from these the noise figure and signal-to-noise ratio for any given signal source. That's all you ever need to know about the noise performance of an amplifier. Basically the process consists of putting known noise signals across the input, then measuring the output noise signal amplitudes within a certain bandwidth. In some cases (e.g., a matched input impedance device such as a radio-frequency amplifier) an oscillator of accurately known and controllable amplitude is substituted as the input signal source.

Later we will discuss the techniques you need to do the output voltage measurement and bandwidth limiting. For now, let's assume you can make rms measurements of the output signal, with a measurement bandwidth of your choice.

the premium low-noise op-amps really are.

□ Low-noise preamps

In addition to the low-noise op-amps, there are some nice low-noise IC *preamplifiers*. Unlike op-amps, these generally have fixed voltage gain, though in some models you can attach an external gain-setting resistor. People sometimes call these "video amplifiers" because they often have bandwidths into the tens of megahertz, though they can be used for low-frequency applications as well. Examples are the Plessey SL561B and several models from Analog Systems. These amplifiers typically have e_n less than $1\text{ nV}/\sqrt{\text{Hz}}$, achieved (at the expense of high input noise *current*, i_n) by running the input transistor at relatively high collector current.

□ NOISE MEASUREMENTS AND NOISE SOURCES

It is a relatively straightforward process to determine the equivalent noise voltage and

□ 7.18 Measurement without a noise source

For an amplifier stage made from a FET or transistor and intended for use at low to moderate frequencies, the input impedance is likely to be very high. You want to know e_n and i_n so that you can predict the SNR with a signal source of arbitrary source impedance and signal level, as we discussed earlier. The procedure is simple:

First, determine the amplifier's voltage gain G_V by actual measurement with a signal in the frequency range of interest. The amplitude should be large enough to override amplifier noise, but not so large as to cause amplifier saturation.

Second, short the input and measure the rms noise output voltage, e_s . From this you get the input noise voltage per root hertz from

$$e_n = \frac{e_s}{G_V B^{\frac{1}{2}}} \quad \text{V}/\text{Hz}^{\frac{1}{2}}$$

where B is the bandwidth of the measurement (see Section 7.21).

Third, put a resistor R across the input, and measure the new rms noise output

voltage, e_r . The resistor value should be large enough to add significant amounts of current noise, but not so large that the input impedance of the amplifier begins to dominate. (If this is impractical, you can leave the input open and use the amplifier's input impedance as R .) The output you measure is just

$$e_r^2 = [e_n^2 + 4kTR + (i_n R)^2] BG_V^2$$

from which you can determine i_n to be

$$i_n = \frac{1}{R_s} \left[\frac{e_r^2}{BG_V^2} - (e_n^2 + 4kTR) \right]^{\frac{1}{2}}$$

With some luck, only the first term in the square root will matter (i.e., if current noise dominates both amplifier voltage noise and source resistor Johnson noise).

Now you can determine the SNR for a signal V_s of source impedance R_s , namely

$$\begin{aligned} \text{SNR} &= 10 \log_{10} \left(\frac{V_s^2}{V_n^2} \right) \\ &= 10 \log_{10} \left[\frac{V_s^2}{[e_n^2 + (i_n R_s)^2 + 4kTR_s]B} \right] \end{aligned}$$

where the numerator is the signal voltage (presumed to lie within the bandwidth B) and the terms in the denominator are the amplifier noise voltage, amplifier noise current applied to R_s , and Johnson noise in R_s . Note that increasing the amplifier bandwidth beyond what is necessary to pass the signal V_s only decreases the final SNR. However, if V_s is broadband (e.g., a noise signal itself), the final SNR is independent of amplifier bandwidth. In many cases the noise will be dominated by one of the terms in the preceding equation.

□ 7.19 Measurement with noise source

The preceding technique of measuring the noise performance of an amplifier has the advantage that you don't need an accurate and adjustable noise source, but it requires an accurate voltmeter and filter, and it

assumes that you know the gain versus frequency of the amplifier, with the actual source resistance applied. An alternative method of noise measurement involves applying broadband noise signals of known amplitude to the amplifier's input and observing the relative increase of output noise voltage. Although this technique requires an accurately calibrated noise source, it makes no assumptions about the properties of the amplifier, since it measures the noise properties right at the point of interest, at the input.

Again, it is relatively straightforward to make the requisite measurements. You connect the noise generator to the amplifier's input, making sure that its source impedance R_g equals the source impedance of the signal you ultimately plan to use with the amplifier. You first note the amplifier's output rms noise voltage, with the noise source attenuated to zero output signal. Then you increase the noise source rms amplitude V_g until the amplifier's output rises 3dB (a factor of 1.414 in rms output voltage). The amplifier's input noise voltage in the measurement bandwidth, for this source impedance, equals this value of added signal. The amplifier therefore has a noise figure

$$\text{NF} = 10 \log_{10} \left(\frac{V_g^2}{4kTR_g} \right)$$

From this you can figure out the SNR for a signal of any amplitude with this same source impedance, using the formula from Section 7.12

$$\text{SNR} = 10 \log_{10} \left(\frac{V_s^2}{4kTR_s} \right) - \text{NF}(R_s) \text{ dB}$$

There are nice calibrated noise sources available, most of which provide means for attenuation to precise levels in the microvolt range. Note: Once again, the preceding formulas assume $R_{in} \gg R_s$. If, on the other hand, the noise-figure measurement is made with a *matched* signal source, i.e., if $R_s = Z_{in}$, then

omit the factors “4” in the preceding expressions.

Note that this technique does not tell you e_n and i_n directly, just the appropriate combination for a source of impedance equal to the driving impedance you used in the measurement. Of course, by making several such measurements with different noise source impedances, you could infer the values of e_n and i_n .

A nice variation on this technique is to use resistor Johnson noise as the “noise source.” This is a favorite technique used by designers of very low noise radiofrequency amplifiers (in which, incidentally, the signal source impedance is usually 50Ω and matches the amplifier’s input impedance). It is usually done the following way: A dewar of liquid nitrogen holds a 50 ohm “termination” (a fancy name for a well-designed resistor that has negligible inductance or capacitance) at the temperature of boiling nitrogen, 77°K ; a second 50 ohm termination is kept at room temperature. The amplifier’s input is connected alternately to the two resistors (usually with a high-quality coax relay), while the output noise power (at some center frequency, with some measurement bandwidth) is measured with an RF power meter. Call the results of the two measurements P_C and P_H , the output noise power corresponding to cold and hot source resistors, respectively. It is then easy to show that the amplifier’s noise temperature, at the frequency of the measurement, is just

$$T_n = \frac{T_H - YT_C}{Y - 1}$$

where $Y = P_H/P_C$, the ratio of noise powers. Noise figure is then given by the formula of Section 7.12, namely

$$\text{NF(dB)} = 10 \log_{10} \left(\frac{T_n}{290} + 1 \right)$$

EXERCISE 7.6

Derive the foregoing expression for noise temperature. Hint: Begin by noting that $P_H = \alpha(T_n + T_H)$ and $P_C = \alpha(T_n + T_C)$, where α is a constant that will shortly disappear. Then note that the noise contribution of the amplifier, stated as a noise temperature, adds to the noise temperature of the source resistor. Take it from there.

EXERCISE 7.7

Amplifier noise temperature (or noise figure) depends on the value of signal source impedance, R_s . Show that an amplifier characterized by e_n and i_n (as in Fig. 7.46) has minimum noise temperature for a source impedance $R_s = e_n/i_n$. Then show that the noise temperature, for that value of R_s , is given by $T_n = e_n i_n / 2k$.

□ Amplifiers with matched input impedance

This last technique is ideal for noise measurements of amplifiers designed for a matched signal source impedance. The most common examples are in radiofrequency amplifiers or receivers, usually meant to be driven with a signal source impedance of 50 ohms, and which themselves have an input impedance of 50 ohms. We will discuss in Chapter 13 the reasons for this departure from our usual criterion that a signal source should have a small source impedance compared with the load it drives. In this situation e_n and i_n are irrelevant as separate quantities; what matters is the overall noise figure (with matched source) or some specification of SNR with a matched signal source of specified amplitude.

Sometimes the noise performance is explicitly stated in terms of the *narrowband* input signal amplitude required to obtain a certain output SNR. A typical radiofrequency receiver might specify a 10dB SNR with a $0.25\mu\text{V}$ rms input signal and 2kHz receiver bandwidth. In this case the procedure consists of measuring the rms receiver output with the input driven by a

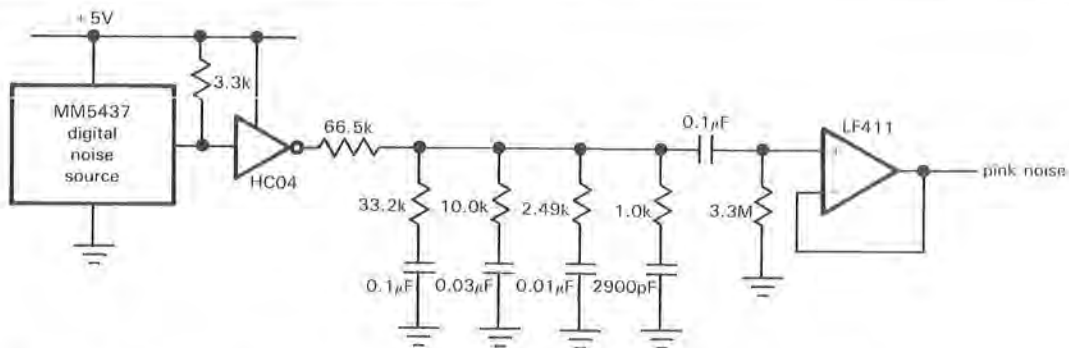


Figure 7.61. Pink noise source (-3dB/octave , $\pm 0.25\text{dB}$ from 10Hz to 40kHz).

matched sine-wave source initially attenuated to zero, then increasing the (sine-wave) input signal until the rms output rises 10dB , in both cases with the receiver bandwidth set to 2kHz . It is important to use a meter that reads true rms voltages for a measurement where noise and signal are combined (more about this later). Note that radiofrequency noise measurements often involve output signals that are in the audiofrequency range.

7.20 Noise and signal sources

Broadband noise can be generated from the effects we discussed earlier, namely Johnson noise and shot noise. The shot noise in a vacuum diode is a classic source of broadband noise that is especially useful because the noise voltage can be predicted exactly. More recently, zener diode noise has been used in noise sources. Both of these extend from dc to very high frequencies, making them useful in audiofrequency and radiofrequency measurements.

An interesting noise source can be made using digital techniques, in particular by connecting long shift registers with their input derived from a modulo-2 addition of several of the later bits (see Section 9.33). The resultant output is a pseudorandom sequence of 1's and 0's that after low-pass filtering generates an analog signal of white spectrum up to the low-pass filter's

break-point, which must be well below the frequency at which the register is shifted. These things can be run at very high frequencies, generating noise up to 100kHz or more. The "noise" has the interesting property that it repeats itself exactly after a time interval that depends on the register length (an n -bit maximal-length register goes through $2^n - 1$ states before repeating). Without much difficulty that time can be made to be very long (months or years), although most often a period of a second is long enough. For example, a 50 -bit register shifted at 10MHz will generate white noise up to 100kHz or so, with a repeat time of 3.6 years. A design for a pseudorandom noise source based on this technique is shown in Section 9.36.

Some noise sources can generate pink noise as well as white noise. Pink noise has equal noise power per *octave*, rather than equal power per hertz. Its power density (power per hertz) drops off at 3dB/octave . Since an RC filter drops off at 6dB/octave , a more complicated filter is necessary to generate a pink spectrum from a white noise input. The circuit shown in Figure 7.61 uses a 23-bit pseudorandom white noise generator chip to generate pink noise, accurate to $\pm 0.25\text{dB}$ from 10Hz to 40kHz .

Versatile signal sources are available with precisely controlled output amplitude

(down to the microvolt range and below) over frequencies from a fraction of a hertz to gigahertz. Some can even be programmed via a digital “bus.” An example is the Hewlett-Packard model 8660 synthesized signal generator, with output frequencies from 0.01 to 110MHz, calibrated amplitudes from 10nV to 1 volt rms, handsome digital display and bus interface, and nifty accessories that extend the frequency range to 2.6GHz and provide modulation and frequency sweeping. This is a bit more than you usually need to do the job.

□ 7.21 Bandwidth limiting and rms voltage measurement

□ Limiting the bandwidth

All the measurements we have been talking about assume that you are looking at the noise output only in a limited frequency band. In a few cases the amplifier may have provision for this, making your job easier. If not, you have to hang some sort of filter on the amplifier output before measuring the output noise voltage.

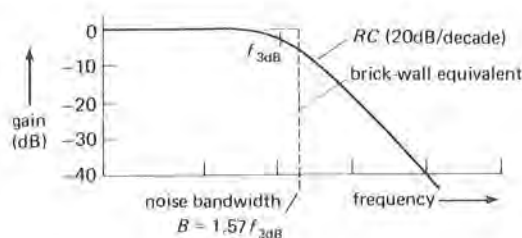


Figure 7.62. Equivalent “brick-wall” noise bandwidth for RC low-pass filter.

The easiest thing to use is a simple RC low-pass filter, with 3dB point set at roughly the bandwidth you want. For accurate noise measurements, you need to know the equivalent “noise bandwidth,” i.e., the width of a perfect “brick-wall” low-pass filter that lets through the same noise voltage (Fig. 7.62). This noise bandwidth is what should be used for B in all the

preceding formulas. It is not terribly difficult to do the mathematics, and you find

$$B = \frac{\pi}{2} f_{3dB} = 1.57 f_{3dB}$$

For a pair of cascaded RCs (buffered so they don’t load each other), the magic formula becomes $B = 1.22 f_{3dB}$. For the Butterworth filters discussed in Section 5.05, the noise bandwidth is

$$B = 1.57 f_{3dB} \quad (1 \text{ pole})$$

$$B = 1.11 f_{3dB} \quad (2 \text{ poles})$$

$$B = 1.05 f_{3dB} \quad (3 \text{ poles})$$

$$B = 1.025 f_{3dB} \quad (4 \text{ poles})$$

If you want to make band-limited measurements up at some center frequency, you can just use a pair of RC filters (Fig. 7.63), in which case the noise bandwidth is as indicated. If you have had experience with contour integration, you may wish to try the following exercise.

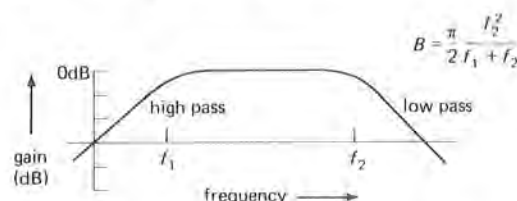


Figure 7.63. Equivalent “brick-wall” noise bandwidth for RC bandpass filter.

EXERCISE 7.8

Optional exercise: Derive the preceding result, beginning with the response functions of RC filters. Assume unit power per hertz input signal, and integrate the output power from zero to infinity. A contour integral then gets you the answer.

Another way to make a bandpass filter for noise measurements is to use an RLC circuit. This is better than a pair of cascaded high-pass and low-pass RC filters if you want your measurement over a bandpass that is narrow compared with the center frequency (i.e., high Q). Figure 7.64 shows both parallel and series RLC circuits and their exact noise bandwidths. In both

cases the resonant frequency is given by $f_0 = 1/2\pi\sqrt{LC}$. You might arrange the bandpass filter circuit as a parallel RLC collector (or drain) load, in which case you use the expression as given. Alternatively, you might interpose the filter as shown in Figure 7.65; for noise bandwidth purposes the circuit is exactly equivalent to the parallel RLC , with $R = R_1 \parallel R_2$.

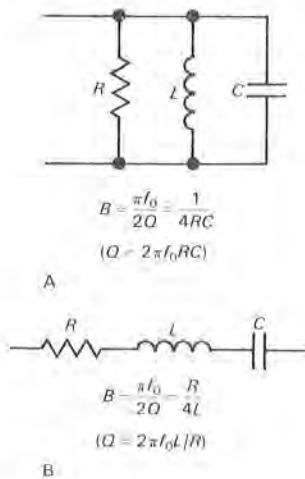


Figure 7.64. Equivalent "brick-wall" noise bandwidth for RLC bandpass filter.

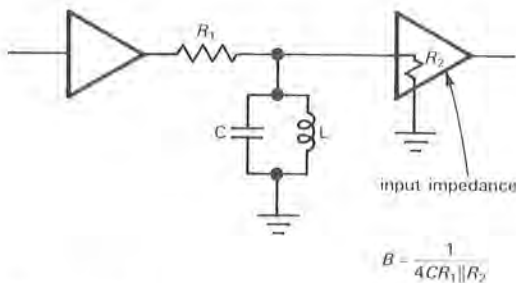


Figure 7.65

□ Measuring the noise voltage

The most accurate way to make output noise measurements is to use a true rms voltmeter. These operate either by measuring the heating produced by the signal waveform (suitably amplified) or by using

an analog squaring circuit followed by averaging. If you use a true rms meter, make sure it has response at the frequencies you are measuring; some of them only go up to a few kilohertz. True rms meters also specify a "crest factor," the ratio of peak voltage to rms that they can handle without great loss of accuracy. For Gaussian noise, a crest factor of 3 to 5 is adequate.

You can use a simple averaging-type ac voltmeter instead, if a true rms meter is unavailable. In that case, the values read off the scale must be corrected. As it turns out, all averaging meters (VOMs, DMMs, etc.) already have their scales adjusted, so what you read isn't actually the *average*, but rather the rms voltage *assuming a sine-wave signal*. For example, if you measure the power-line voltage in the United States, your meter will read something close to 117 volts. That's fine, but if the signal you're reading is Gaussian noise, you have to apply an additional correction. The rule is as follows: To get the rms voltage of Gaussian noise, multiply the "rms" value you read on an averaging ac voltmeter by 1.13 (or add 1dB). Warning: This works fine if the signal you are measuring is pure noise (e.g., the output of an amplifier with a resistor or noise source as input), but it won't give accurate results if the signal consists of a sine wave added to noise.

A third method, not exactly world-famous for its accuracy, consists of looking at the noise waveform on an oscilloscope: The rms voltage is 1/6 to 1/8 of the peak-to-peak value (depending on your subjective reading of the pp amplitude). It isn't very accurate, but at least there's no problem getting enough measurement bandwidth.

7.22 Noise potpourri

Herewith a collection of interesting, and possibly useful, facts.

1. The averaging time required in an indicating device to reduce the fluctuations of

a rectified noise signal to a desired level for a given noise bandwidth is

$$\tau \approx \frac{1600}{B\sigma_2} \text{ seconds}$$

where τ is the required time constant of the indicating device to produce fluctuations of standard deviation σ percent at the output of a linear detector whose input is noise of bandwidth B .

2. For band-limited white noise the expected number of maxima per second is

$$N = \sqrt{\frac{3(f_2^5 - f_1^5)}{5(f_2^3 - f_1^3)}}$$

where f_1 and f_2 are the lower and upper band limits. For $f_1 = 0$, $N = 0.77f_2$; for narrowband noise ($f_1 \approx f_2$), $N \approx (f_1 + f_2)/2$.

3. rms-to-average (i.e., average magnitude) ratios:

$$\text{Gaussian noise: } \text{rms/avg} = \sqrt{\pi/2} = 1.25 \\ = 1.96\text{dB}$$

$$\text{Sine wave: } \text{rms/avg} = \pi/2^{3/2} = 1.11 \\ = 0.91\text{dB}$$

$$\text{Square wave: } \text{rms/avg} = 1 = 0\text{dB}$$

4. Relative occurrence of amplitudes in Gaussian noise. Figure 7.66 gives the fractional time that a given amplitude level is exceeded by a Gaussian noise waveform of amplitude 1 volt rms.

INTERFERENCE: SHIELDING AND GROUNDING

"Noise" in the form of interfering signals, 60Hz pickup, and signal coupling via power supplies and ground paths can turn out to be of far greater practical importance than the intrinsic noise sources we've just discussed. These interfering signals can all be reduced to an insignificant level (unlike thermal noise) with proper layout and construction. In stubborn cases the cure may involve a combination of filtration of input and output lines, careful

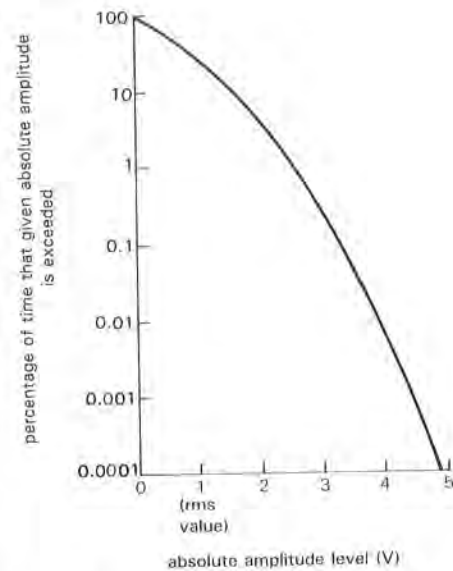


Figure 7.66. Relative occurrence of amplitudes in Gaussian noise.

layout and grounding, and extensive electrostatic and magnetic shielding. In these sections we would like to offer some suggestions that may help illuminate this dark area of the electronic art.

7.23 Interference

Interfering signals can enter an electronic instrument through the power-line inputs or through signal input and output lines. In addition, signals can be capacitively coupled (electrostatic coupling) onto wires in the circuit (the effect is more serious for high-impedance points within the circuit), magnetically coupled to closed loops in the circuit (independent of impedance level), or electromagnetically coupled to wires acting as small antennas for electromagnetic radiation. Any of these can become a mechanism for coupling of signals from one part of a circuit to another. Finally, signal currents from one part of the circuit can couple to other parts through voltage drops on ground lines or power-supply lines.

Eliminating interference

Numerous effective tricks have been evolved to handle most of these commonly occurring interference problems. Keep in mind the fact that these techniques are all aimed at reducing the interfering signal or signals to an acceptable level; they rarely eliminate them altogether. Consequently, it often pays to raise signal levels, just to improve the signal-to-interference ratio. Also, it is important to realize that some environments are much worse than others; an instrument that works just dandy on the bench may perform miserably on location. Some environments worth avoiding are those (a) near a radio or television station (RF interference), (b) near a subway (impulsive interference and power-line garbage), (c) near high-voltage lines (radio interference, frying sounds), (d) near motors and elevators (power-line spikes), (e) in a building with triac lamp and heater controllers (power-line spikes), (f) near equipment with large transformers (magnetic pickup), and (g) near arc welders (unbelievable pickup of all sorts). Herewith a gathering of advice, techniques, and black magic:

Signals coupled through inputs, outputs, and power line

The best bet for power-line noise is to use a combination of RF line filters and transient suppressors on the ac power line. You can achieve 60dB or better attenuation of interference above a few hundred kilohertz this way, as well as effective elimination of damaging spikes.

Inputs and outputs are more difficult, because of impedance levels and the need to couple desired signals that may lie in the frequency range of interference. In devices like audio amplifiers you can use low-pass filters on inputs *and* outputs (much interference from nearby radio stations enters via the speaker wires, acting as antennas). In other situations shielded lines are often

necessary. Low-level signals, particularly at high impedance levels, should always be shielded. So should the instrument cabinet.

Capacitive coupling

Signals within an instrument can get around handsomely via electrostatic coupling: Some point within the instrument has a 10 volt signal jumping around; a high- Z input nearby does some sympathetic jumping, too. The best things to do are to reduce the capacitance between the offending points (move them apart), add shielding (a complete metal enclosure, or even close-knit metal screening, eliminates this form of coupling altogether), move the wires close to a ground plane (which “swallows” the electrostatic fringing fields, reducing coupling enormously), and lower the impedance levels at susceptible points, if possible. Op-amp outputs don’t pick up interference easily, whereas inputs do. More on this later.

Magnetic coupling

Unfortunately, low-frequency magnetic fields are not significantly reduced by metal enclosures. A turntable, microphone, tape recorder, or other sensitive circuit placed in close proximity to an instrument with a large power transformer will display astounding amounts of 60Hz pickup. The best therapy here is to avoid large enclosed areas within circuit paths and try to keep the circuit from closing around in a loop. Twisted pairs of wires are quite effective in reducing magnetic pickup, because the enclosed area is small, and the signals induced in successive twists cancel.

When dealing with very low level signals, or devices particularly susceptible to magnetic pickup (tape heads, inductors, wire-wound resistors), it may be desirable to use magnetic shielding. “Mu-metal shielding” is available in preformed pieces

and flexible sheets. If the ambient magnetic field is large, it is best to use shielding of high permeability (high μ) on the inside, surrounded by an outer shield of lower permeability (which can be ordinary iron, or low- μ shielding material), to prevent magnetic saturation in the inner shield. Of course, moving the offending source of magnetic field is often a simpler solution. It may be necessary to exile large power transformers to the hinterlands, so to speak. Toroidal transformers have smaller fringing fields than the standard frame types.

Radiofrequency coupling

RF pickup can be particularly insidious, because innocent-looking parts of the circuit can act as resonant circuits, displaying enormous effective cross section for pickup. Aside from overall shielding, it is best to keep leads short and avoid loops that can resonate. Ferrite beads may help, if the problem involves very high frequencies. A classic situation is the use of a pair of bypass capacitors (one tantalum, one disc ceramic), often recommended to improve bypassing. The pair can form a lovely parasitic tuned circuit somewhere in the HF to VHF region (tens to hundreds of megahertz), with self-oscillations!

7.24 Signal grounds

Ground leads and shields can cause plenty of trouble, and there is a lot of misunderstanding on this subject. The problem, in a nutshell, is that currents you forgot about flowing through a ground line can generate a signal seen by another part of the circuit sharing the same ground. The technique of a ground "mecca" (a common point in the circuit to which all ground connections are tied) is often seen, but it's a crutch; with a little understanding of the problem you can handle most situations intelligently.

Common grounding blunders

Figure 7.67 shows a common situation. Here a low-level amplifier and a high-current driver are in the same instrument. The first circuit is done correctly: Both amplifiers tie to the supply voltages at the regulator (right at the sensing leads), so IR drops along the leads to the power stage don't appear on the low-level amplifier's supply voltages. In addition, the load current returning to ground does not appear at the low-level input; no current flows from the ground side of the low-level amplifier's input to the circuit mecca (which might be the connection to the case near the BNC input connector).

In the second circuit there are two blunders. Supply voltage fluctuations caused by load currents at the high-level stage are impressed on the low-level supply voltages. Unless the input stage has very good supply rejection, this can lead to oscillations. Even worse, the load current returning to the supply makes the case "ground" fluctuate with respect to power-supply ground. The input stage ties to this fluctuating ground, a very bad idea. The general idea is to look at where the large signal currents are flowing and make sure their IR drops don't wind up at the input. In some cases it may be a good idea to decouple the supply voltages to the low-level stages with a small RC network (Fig. 7.68). In stubborn cases of supply coupling it may pay to put a zener or 3-terminal regulator on the low-level-stage supply for additional decoupling.

□ 7.25 Grounding between instruments

The idea of a controlled ground point within one instrument is fine, but what do you do when a signal has to go from one instrument to another, each with its own idea of "ground"? Some suggestions follow.

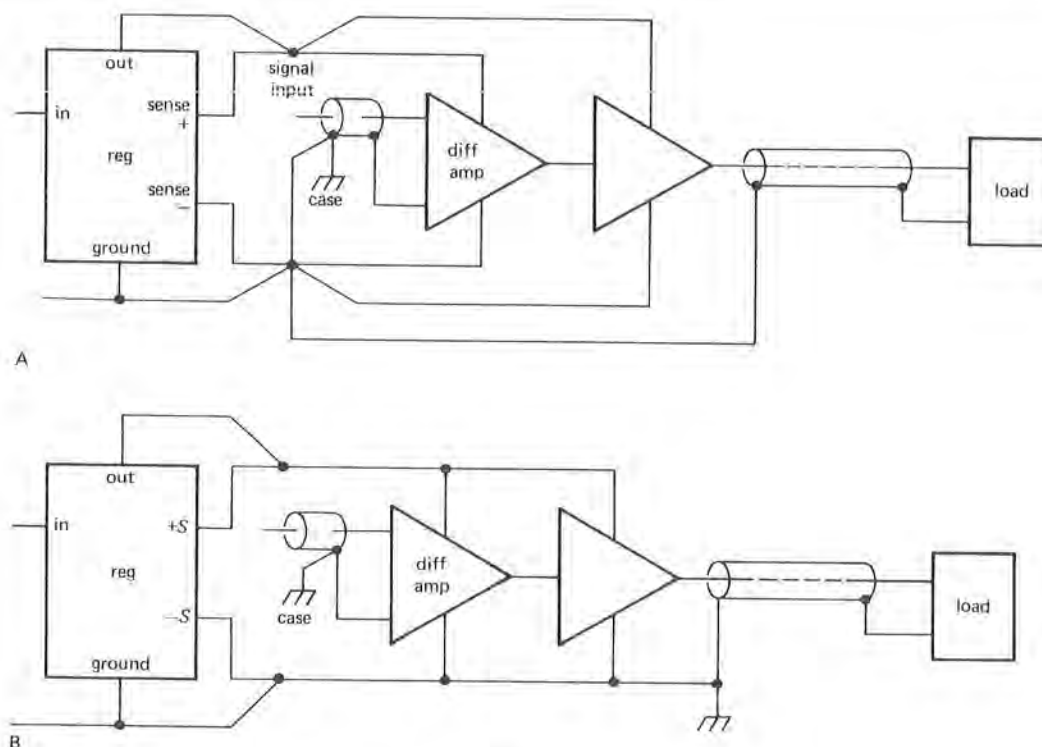


Figure 7.67. Ground paths for low-level signals.
A. Right
B. Wrong

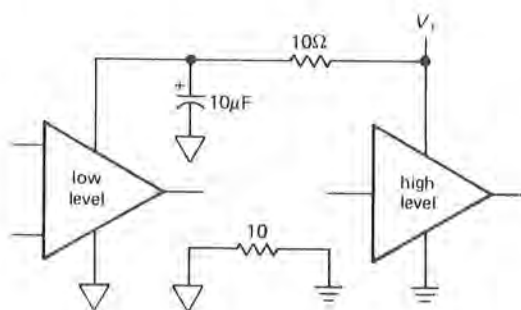


Figure 7.68

□ High-level signals

If the signals are several volts, or large logic swings, just tie things together and forget about it (Fig. 7.69). The voltage source shown between the two grounds represents the variations in local grounds you'll find on different power-line outlets in the same

room or (worse) in different rooms or buildings. It consists of some 60Hz voltage, harmonics of the line frequency, some radiofrequency signals (the power line makes a good antenna), and assorted spikes and other garbage. If your signals are large enough, you can live with this.

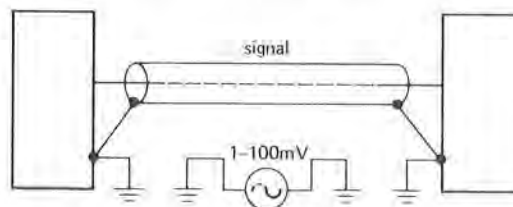


Figure 7.69

□ Small signals and long wires

For small signals this situation is intolerable, and you will have to go to some effort

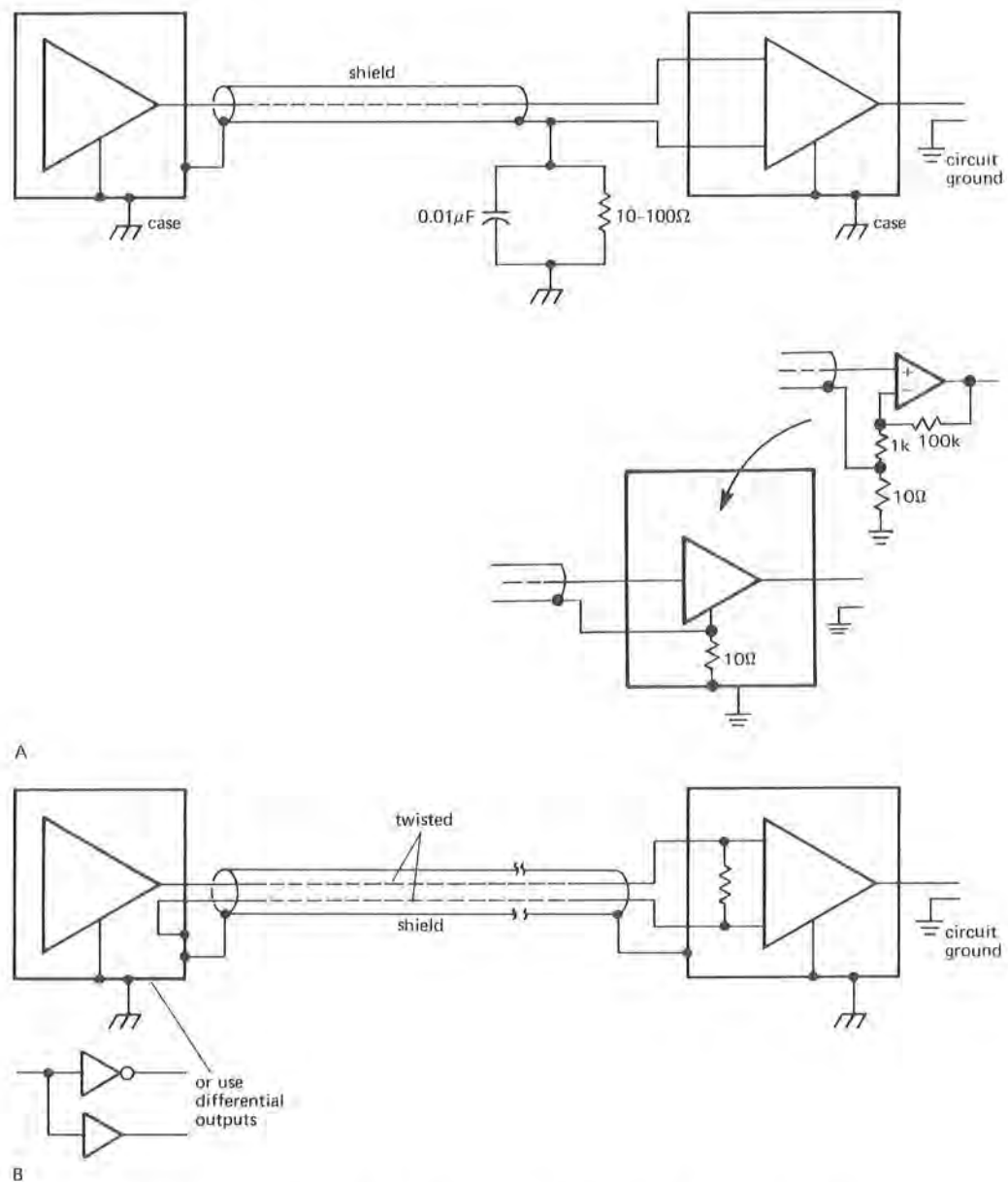


Figure 7.70. Ground connections for low-level signals through shielded cables.

to remedy the situation. Figure 7.70 shows some ideas. In the first circuit, a coaxial shielded cable is tied to the case and circuit ground at the driving end, but it is kept isolated from the case at the receiving end

(use a Bendix 4890-1 or Amphenol 31-010 insulated BNC connector). A differential amplifier is used to buffer the input signal, thus ignoring the small amount of "ground signal" appearing on the shield. A small

resistor and bypass capacitor to ground is a good idea to limit ground swing and prevent damage to the input stage. The alternate receiver circuit in Figure 7.70 shows the use of a “pseudodifferential” input connection for a single-ended amplifier stage (which might, for example, be a standard non-inverting op-amp connection, as indicated). The 10 ohm resistor between amplifier common and circuit ground is large enough to let the signal source’s reference ground set the potential at that point, since it is much larger than the impedance of the source’s ground. Any noise present at that node, of course, appears also at the output. However, this becomes unimportant if the stage has sufficiently high voltage gain, G_V , since the ratio of desired signal to ground noise is reduced by G_V . Thus, although this circuit isn’t truly differential (with infinite CMRR), it works well enough (with effective $\text{CMRR} = G_V$). This pseudo-differential ground-sensing trick can be used also for low-level signals *within* an instrument, when ground noise is a problem.

In the second circuit, a shielded twisted pair is used, with the shield connected to the case at both ends. Since no signal travels on the shield, this is harmless. A differential amplifier is used as before on the receiving end. If logic signals are being transmitted, it is a good idea to send a differential signal (the signal and its inverted form), as indicated. Ordinary differential amplifiers can be used as input stages, or if the ground interference is severe, special “isolated amplifiers” are available from manufacturers like Analog Devices and Burr-Brown. The latter permit kilovolts of common-mode signals. So do opto-isolator modules, a handy solution for digital signals in some situations.

At radiofrequencies, transformer coupling offers a convenient way of removing common-mode signal at the receiving end; this also makes it easy to generate a differential bipolarity signal at the driving end.

Transformers are popular in audio applications as well, although they tend to be bulky and lead to some signal degradation.

For very long cable runs (measured in miles) it is useful to prevent large ground currents flowing in the shield at radiofrequencies. Figure 7.71 suggests a method. As before, a differential amplifier looks at the twisted pair, ignoring the voltage on the shield. By tying the shield to the case through a small inductor, the dc voltage is kept small while preventing large radiofrequency currents. This circuit also shows protection circuitry to prevent common-mode excursions beyond ± 10 volts.

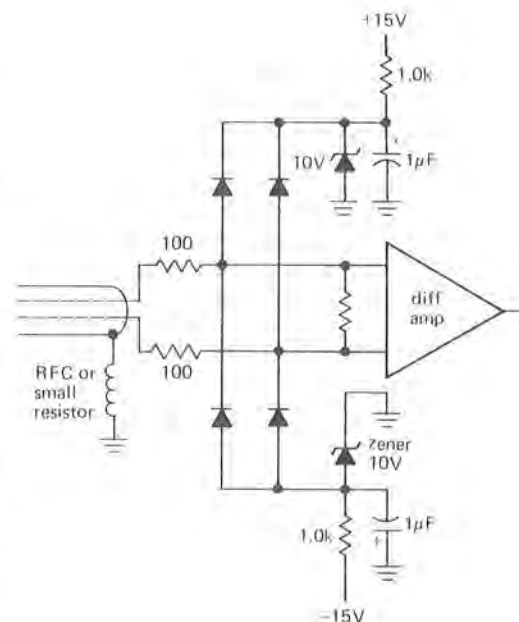


Figure 7.71. Input-protection circuits for use with very long lines.

Figure 7.72 shows a nice scheme to save wires in a multiwire cable in which the common-mode pickup has to be eliminated. Since all the signals suffer the same common-mode pickup, a single wire tied to ground at the sending end serves to cancel the common-mode signals on each of the n signal lines. Just buffer its signal

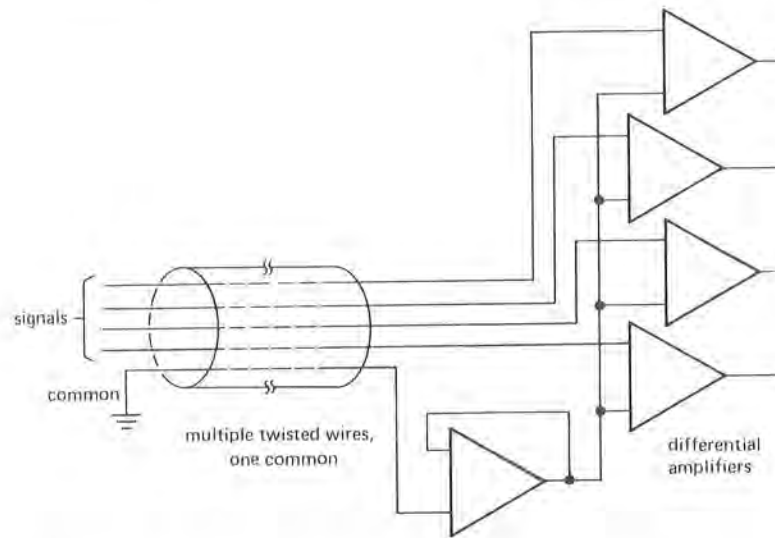


Figure 7.72. Common-mode interference rejection with long multiwire cables.

(with respect to ground at the receiving end), and use it as the comparison input for each of n differential amplifiers looking at the other signal lines.

The preceding schemes work well to eliminate common-mode interference at low to moderate frequencies, but they can be ineffective against radiofrequency interference, owing to poor common-mode rejection in the receiving differential amplifier.

One possibility here is to wrap the whole cable around a ferrite toroid (Fig. 7.73). That increases the series inductance of the whole cable, raising the impedance to common-mode signals of high frequency and making it easy to bypass them at the far end with a pair of small bypass capacitors to ground. The equivalent circuit shows why this works without attenuating the differential signal: You have a series inductance inserted into both signal lines and the shield, but since they form a tightly coupled transformer of unit turns ratio, the differential signal is unaffected. This is actually a "1:1 transmission-line transformer," as discussed in Section 13.10.

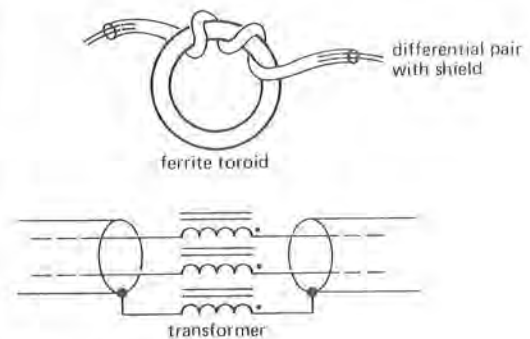


Figure 7.73

□ Floating signal sources

The same sort of disagreement about the voltage of "ground" at separated locations enters in an even more serious way at low-level inputs, just because the signals are so small. An example is a magnetic tape head or other signal transducer that requires a shielded signal line. If you ground the shield at both ends, differences in ground potential will appear as signal at the amplifier input. The best approach is to lift the shield off ground at the transducer (Fig. 7.74).

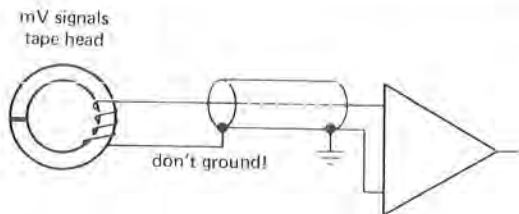


Figure 7.74

Isolation amplifiers

Another solution to serious ground-contention problems is the use of an "isolation amplifier." Isolation amplifiers (iso-amps) are commercial devices intended for coupling an analog signal (with bandwidth clear down to dc) from a circuit with one ground reference to another circuit with a completely different ground (Fig. 7.75). In fact, in some bizarre situations the "grounds" can differ by many kilovolts! Isolation amplifiers are mandatory for medical electronics in which electrodes are applied to human subjects, in order to isolate completely those connections from any instrument circuits powered directly from the ac power lines. Currently available isolation amplifiers use one of three methods:

1. Transformer isolation of a high-

frequency carrier signal, which is either frequency-modulated or pulse-width-modulated with the relatively low bandwidth signal (dc to 10kHz or so) to be isolated (Fig. 7.76). This method is used in all of the isolation amplifiers from Analog Devices, as well as some units from Burr-Brown. Transformer-isolated iso-amps have the convenient feature of requiring dc power only on one side; they all include a transformer-coupled dc-to-dc converter in the package. Transformer-coupled iso-amps can isolate up to 3.5kV and have typical bandwidths of 2kHz, though some units go to 20kHz.

2. Optically coupled signal transmission via an LED at the sending end and photodiode at the receiving end. This technique is typified by the ISO100 from Burr-Brown. No high-frequency carrier is needed, since signals all the way to dc can be transmitted optically. However, to achieve good linearity, Burr-Brown uses a cute trick: A second matched photodiode at the transmitting side receives light from the LED, in a feedback arrangement that cancels nonlinearities in both LED and photodiode; see Figure 7.77. The ISO100 requires power supplies at both ends, isolates to 750 volts, and has 60kHz bandwidth.

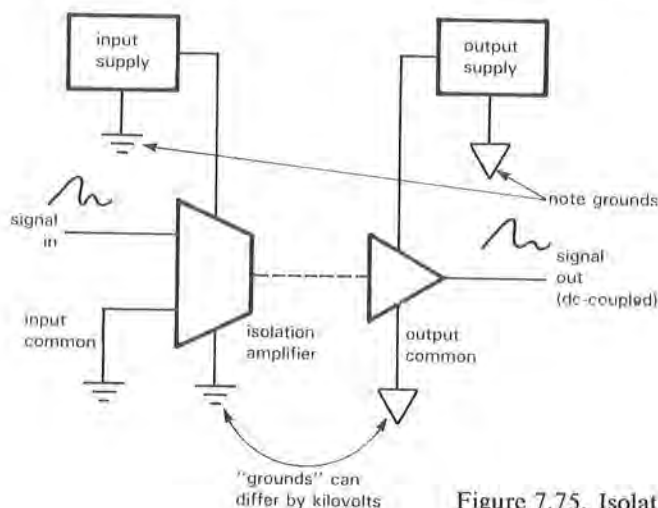


Figure 7.75. Isolation amplifier concept.

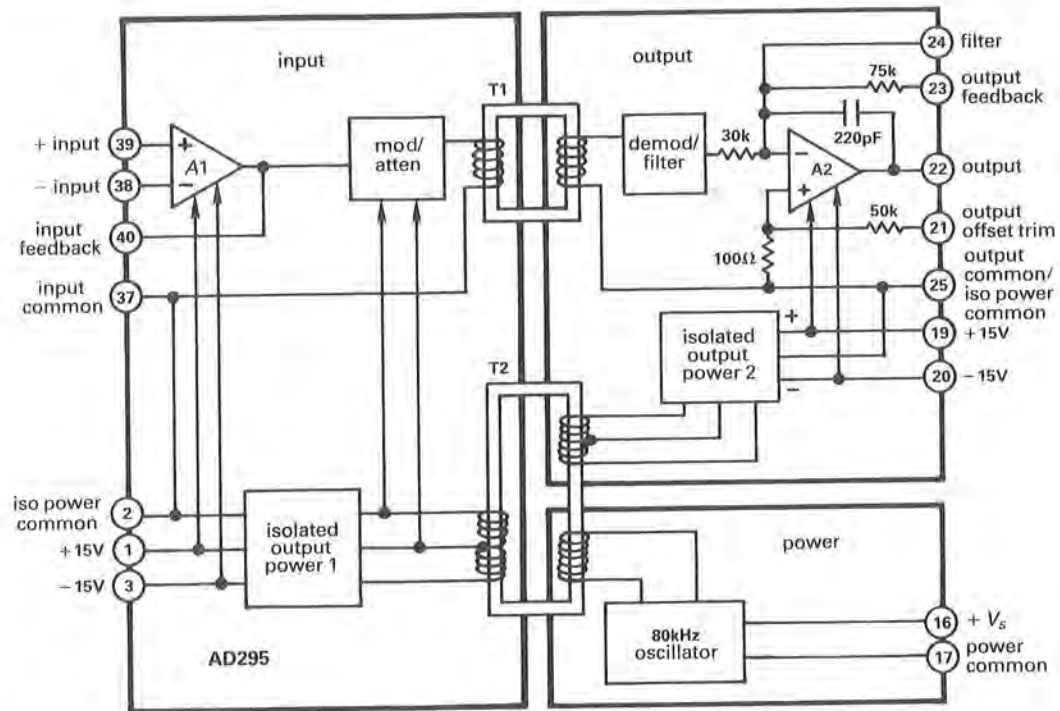


Figure 7.76. AD295 transformer-coupled isolation amplifier. (Courtesy of Analog Devices.)

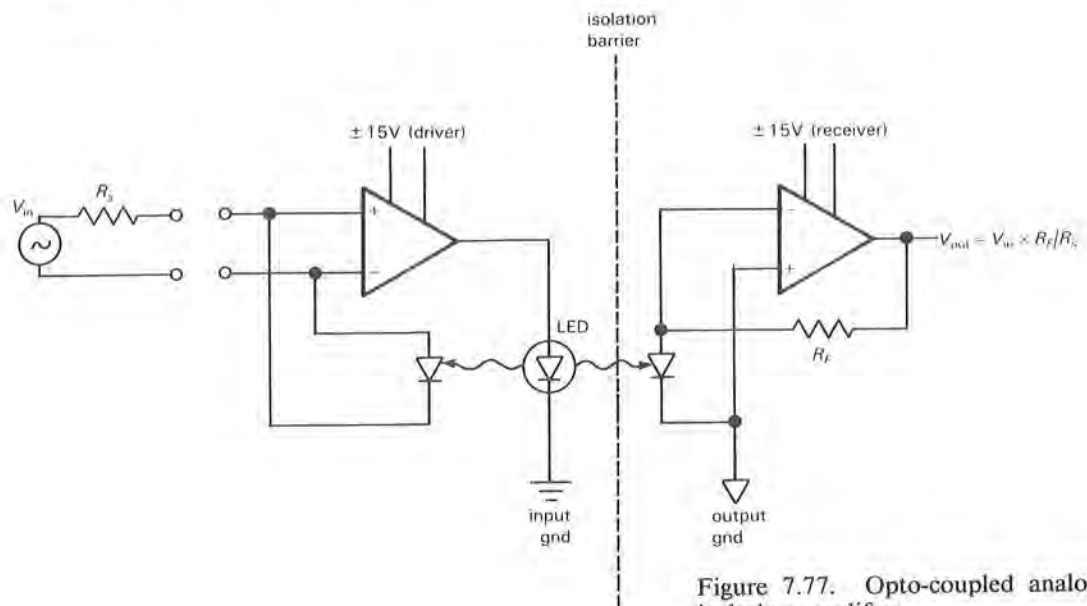


Figure 7.77. Opto-coupled analog isolation amplifier.

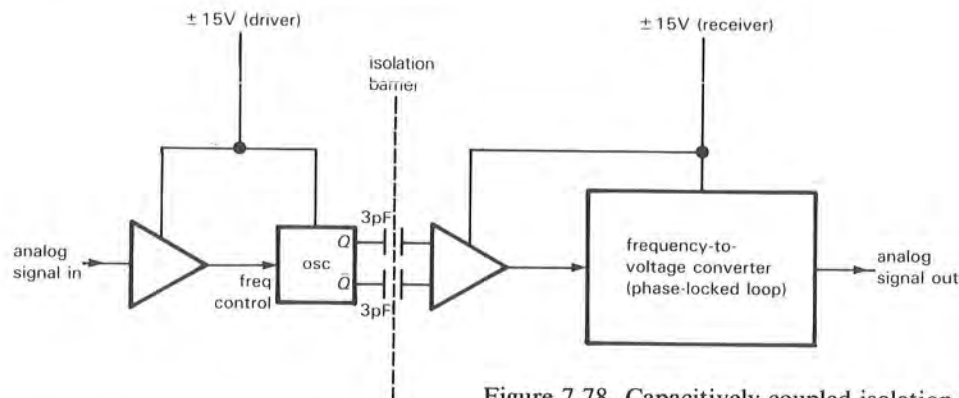


Figure 7.78. Capacitively coupled isolation amplifier.

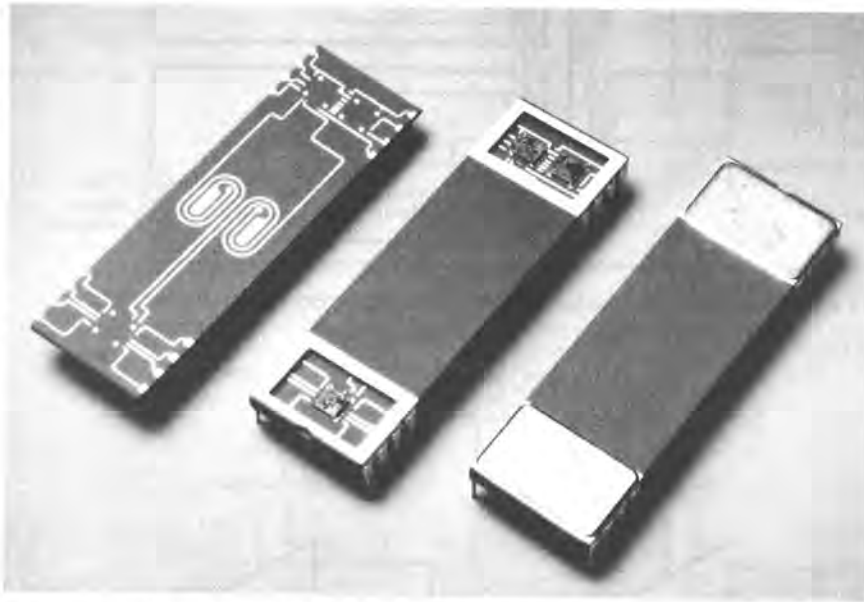


Figure 7.79. Burr-Brown ISO106 isolation amplifier. (Courtesy of Burr-Brown Corporation.)

3. Capacitively coupled isolation of a high-frequency carrier signal, which is frequency-modulated with the signal to be isolated (Fig. 7.78). This technique is typified by the ISO102, ISO106, and ISO122 from Burr-Brown (Fig. 7.79). There is no feedback, as with transformer isolation, but for most models you need power supplies at both ends. This usually isn't a problem, since you are likely to have electronics at both ends, generating and using the signal. If not, you can get an isolated

dc-dc converter to use with the iso-amp. The ISO106 isolates to 3.5kV and has 70kHz bandwidth.

These isolation amplifiers are all intended for *analog* signals, of modest bandwidth; they cost from \$25 to \$100 each. The same sorts of ground problems can arise in digital electronics, where the solution is simple and effective: Optically coupled isolators ("opto-isolators") are available, with plenty of bandwidth (10MHz or more), isolation of several

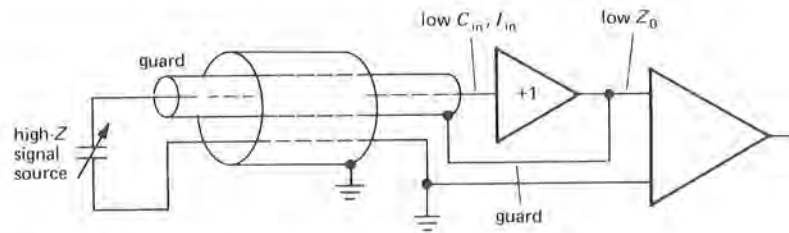


Figure 7.80. Using a guard to raise input impedance.

kilovolts, and low cost (a dollar or two). We'll see them in Chapter 9.

remote amplifier. We will discuss signal guards in Section 15.08 in connection with high-impedance microelectrodes.

□ Signal guarding

A closely related issue is *signal guarding*, an elegant technique to reduce the effects of input capacitance and leakage for small signals at high impedance levels. You may be dealing with signals from a microelectrode or a capacitive transducer, with source impedances of hundreds of megohms. Just a few picofarads input capacitance can form a low-pass filter, with rolloffs beginning at a few hertz! In addition, the effects of insulation resistance in the connecting cables can easily degrade the performance of an ultra-low input current amplifier (bias currents less than a picoamp) by orders of magnitude. The solution to both these problems is a *guard electrode* (Fig. 7.80).

A follower bootstraps the inner shield, effectively eliminating leakage current and capacitive attenuation by keeping zero voltage difference between the signal and its surrounding. An outer grounded shield is a good idea, to keep interference off the guard electrode; the follower has no trouble driving that capacitance and leakage, of course, since it has low output impedance.

You shouldn't use these tricks more than you need to; it would be a good idea to put the follower as close to the signal source as possible, guarding only the short section of cable connecting them. Ordinary shielded cable can then carry the low-impedance output signal out to the

□ Coupling to outputs

Ordinarily the output impedance of an op-amp is low enough that you don't have to worry about capacitive signal coupling. In the case of high-frequency or fast-switching interference, however, you have just cause for alarm, particularly if the desired output signal involves some degree of precision. Consider the example in Figure 7.81. A precision signal is buffered by an op-amp and passes through a region containing digital logic signals jumping around with slew rates of 0.5V/ns. The op-amp's closed-loop output impedance rises with frequency, typically reaching values of 10 to 100 ohms or more at 1MHz (see Section 7.07). How large a coupling capacitance is permissible, keeping coupled interference less than the analog signal's resolution of 0.1mV? The surprising answer is 0.02pF.

There are some solutions. The best thing is to keep your small analog waveforms out of the reach of fast-switching signals. A moderate bypass capacitor across the op-amp's output (with perhaps a small series resistor, to maintain op-amp stability) will help, although it degrades the slew rate. You can think of the action of this capacitor as lowering the frequency of the coupled charge bundles to the point where the op-amp's feedback can swallow them. A

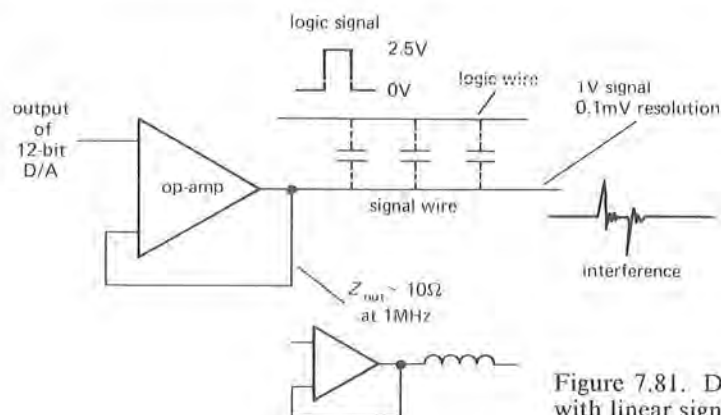


Figure 7.81. Digital cross-coupling interference with linear signals.

few hundred picofarads to ground will adequately stiffen the analog signal at high frequencies (think of it as a capacitive voltage divider). Another possibility is to use a low-impedance buffer, such as the LT1010, or a power op-amp such as the LM675. Don't neglect the opportunity to use shielding, twisted pairs, and proximity to ground planes to reduce coupling.

SELF-EXPLANATORY CIRCUITS

7.26 Circuit ideas

Figure 7.82 presents some circuit ideas relevant to the subjects of this chapter.

ADDITIONAL EXERCISES

(1) Prove that $\text{SNR} = 10 \log_{10}(v_s^2 / 4kTR_s) - \text{NF(dB)}$ (at R_s).

(2) A $10\mu\text{V}$ (rms) sine wave at 100Hz is in series with a 1M resistor at room temperature. What is the SNR of the resultant signal (a) in a 10Hz band centered at 100Hz and (b) in a 1MHz band going from dc to 1MHz?

(3) A transistor amplifier using a 2N5087 is operated at $100\mu\text{A}$ collector current and is driven by a signal source of impedance 2000 ohms. (a) Find the noise figure at 100Hz, 1kHz, and 10kHz. (b) Find the SNR (at each of listed frequencies) for an

input signal of 50nV (rms) and an amplifier bandwidth of 10Hz.

(4) Measurements are made on a commercial amplifier (with $Z_{in} = 1\text{M}$) in order to determine its equivalent input noise e_n and i_n at 1kHz. The amplifier's output is passed through a sharp-skirted filter of bandwidth 100Hz: A $10\mu\text{V}$ input signal results in a 0.1 volt output. At this level the amplifier's noise contribution is negligible. With the input shorted, the noise output is 0.4mV rms. With the input open, the noise output rises to 50mV rms. (a) Find e_n and i_n for this amplifier at 1kHz. (b) Find the noise figure of this amplifier at 1kHz for source resistances of 100 ohms, 10k, and 100k.

(5) Noise measurements are made on an amplifier using a calibrated noise source whose output impedance is 50 ohms. The generator output must be raised to $2\text{nV/Hz}^{1/2}$ in order to double the output noise power of the amplifier. What is the amplifier's noise figure for a source impedance of 50 ohms?

(6) The output noise voltage of a white noise generator is measured with the circuit shown in Figure 7.83. At a particular setting of the noise generator output level, the ac voltmeter reads 1.5 volts "rms." What is the noise generator's output noise density (i.e., rms volts per root hertz)?



Circuit ideas

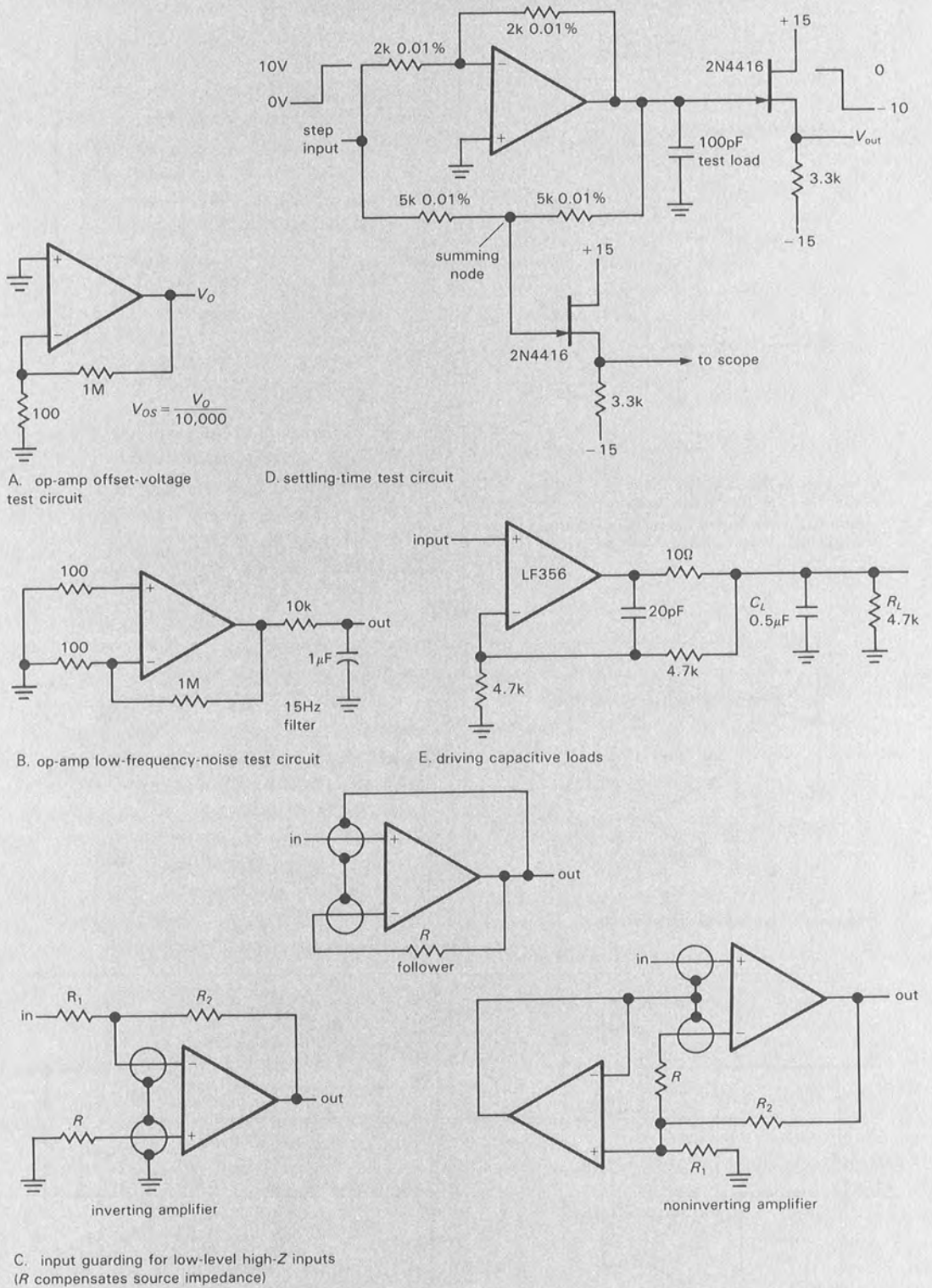
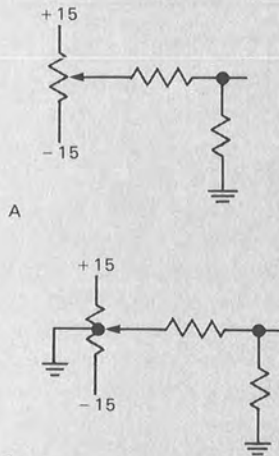


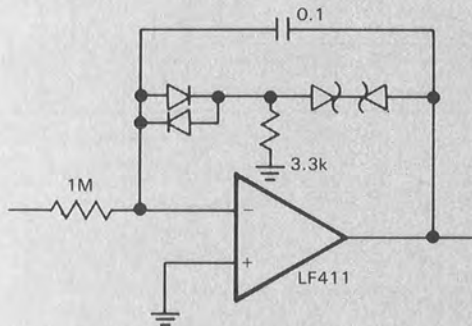
Figure 7.82



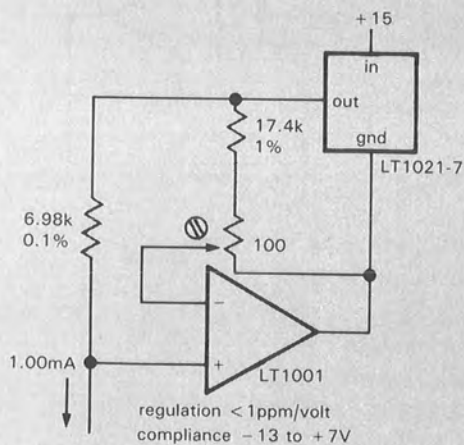
Circuit ideas (cont.)



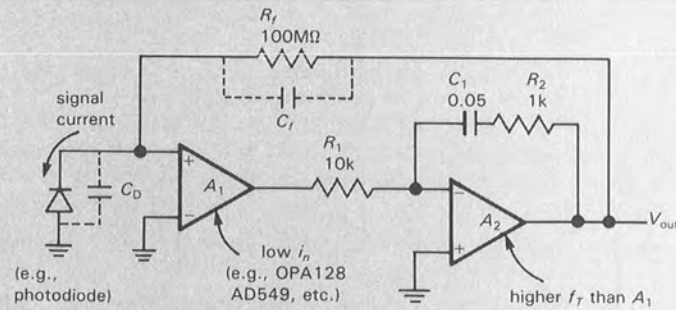
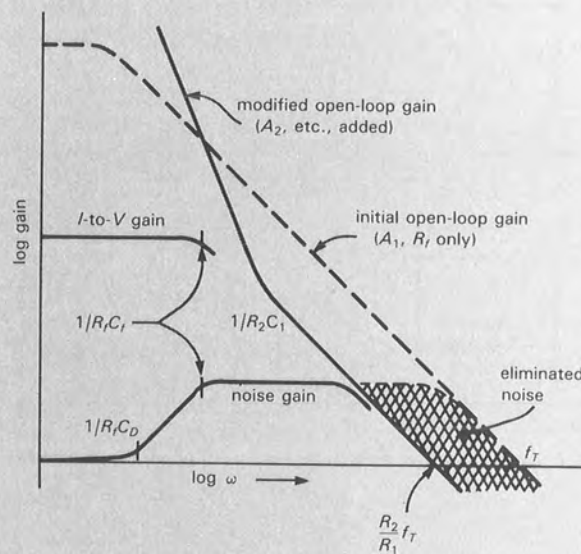
F. A. traditional offset trimming circuit
B. with improved trimpot (available from Bourns), circuit has less dependence on unbalanced supply voltage changes



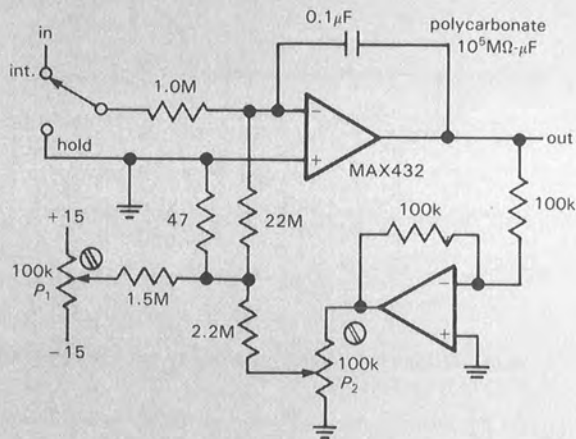
G. integrator with low-leakage zener clamp



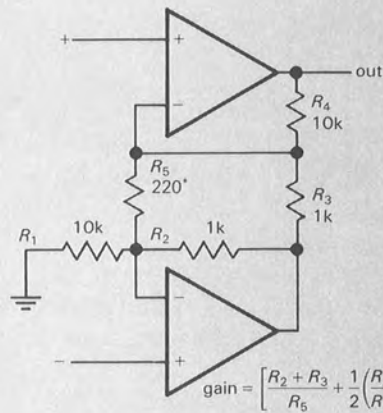
H. ultraprecise current source



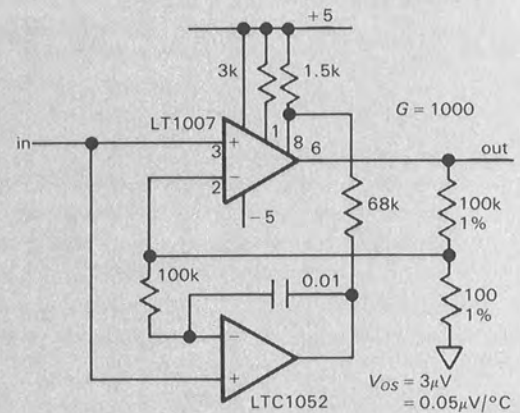
I. reducing "noise gain" in low-level transimpedance amplifier (from Burr-Brown App. Note)



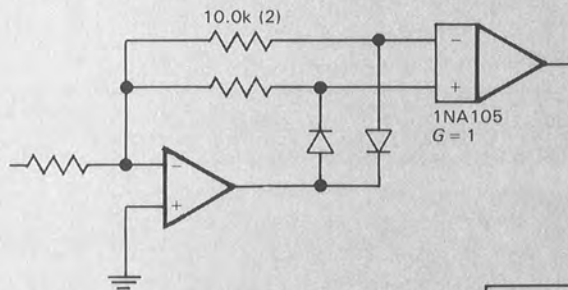
J. precision integrator with capacitor leakage compensation; use P_1 to null hold drift while output is near zero and P_2 while output is near +10V



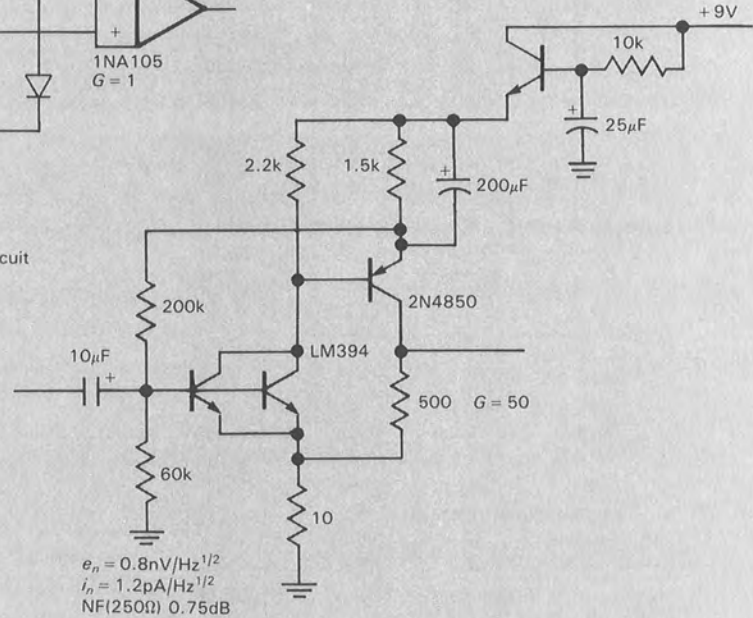
*select gain with changes in R_5
K. instrumentation amplifier



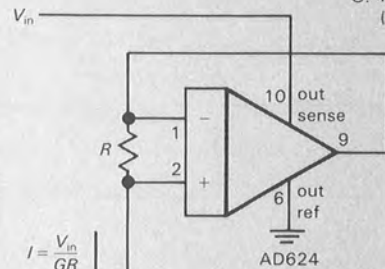
N. low-noise, low-drift amplifier; noise 60nV p-p, 0.1 to 10Hz, rather than the larger value of 1.5μV p-p for the chopper amplifier



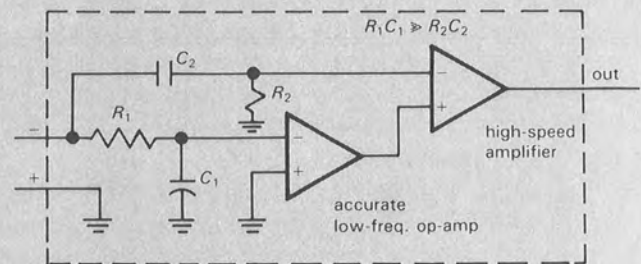
L. active rectifier with symmetrical circuit



O. low-noise preamplifier, intended for $R_S < 600\Omega$
(from NSC design by Bob Widlar)



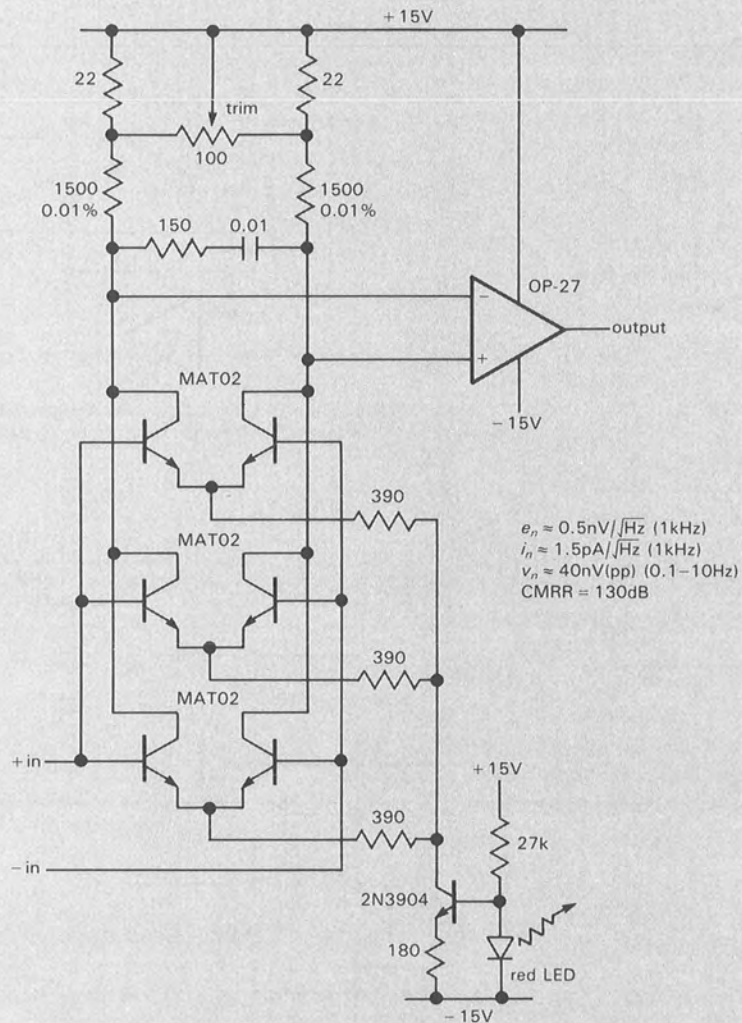
M. programmable current source using instrumentation amplifier



P. dc-accurate high-speed or high-power op-amp



Circuit ideas (cont.)



Q. ultra-low-noise op-amp (from PMI App. Note 102)

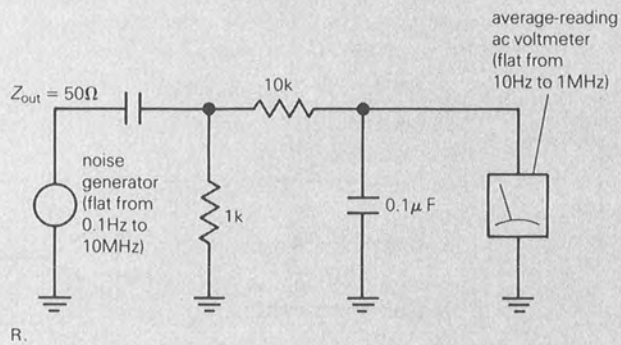


Figure 7.83

DIGITAL ELECTRONICS

CHAPTER 8

BASIC LOGIC CONCEPTS

8.01 Digital versus analog

Thus far we have been dealing mainly with circuits in which the input and output voltages have varied over a range of values: *RC* circuits, amplifiers, integrators, rectifiers, op-amps, etc. This is natural when dealing with signals that are continuous (e.g., audio signals) or continuously varying voltages from measuring instruments (e.g., temperature-reading or light-detecting devices, or biological or chemical probes).

However, there are instances in which the input signal is naturally discrete in form, e.g., pulses from a particle detector, or “bits” of data from a switch, keyboard, or computer. In such cases the use of digital electronics (circuits that deal with data made of 1’s and 0’s) is natural and convenient. Furthermore, it is often desirable to convert continuous (analog) data to digital form, and vice versa (using D/A and A/D converters), in order to perform calculations on the data with a calculator or computer or to store large quantities of data

as numbers. In a typical situation a microprocessor or computer might monitor signals from an experiment or industrial process, control the experimental parameters on the basis of the data obtained, and store for future use the results collected or computed while the experiment is running.

Another interesting example of the power of digital techniques is the transmission of analog signals without degradation by noise: An audio or video signal, for instance, picks up “noise” while being transmitted by cable or radio that cannot be removed. If, instead, the signal is converted to a series of numbers representing its amplitude at successive instants of time, and these numbers are transmitted as digital signals, the analog signal reconstruction at the receiving end (done with D/A converters) will be without error, providing the noise level on the transmission channel isn’t high enough to prevent accurate recognition of 1’s and 0’s. This technique, known as PCM (pulse-code modulation), is particularly attractive where a signal must pass through a series of “repeaters,” as in the case of a transcontinental telephone call, since

digital regeneration at each stage guarantees noiseless transmission. The information and pictures sent back by recent deep space probes were done with PCM. Digital audio is now commonplace in the home, in the form of 12cm optical “compact discs” (CDs); these store a piece of music in the form of a stereo pair of 16-bit numbers every 23 microseconds, 6 billion bits of information in all.

In fact, digital hardware has become so powerful that tasks that seem well suited to analog techniques are often better solved with digital methods. As an example, an analog temperature meter might incorporate a microprocessor and memory in order to improve accuracy by compensating the instrument’s departure from perfect linearity. Because of the wide availability of microprocessors, such applications are becoming commonplace. Rather than attempt to enumerate what can be done with digital electronics, let’s just start learning about it. Applications will emerge naturally as we go along.

8.02 Logic states

By “digital electronics” we mean circuits in which there are only two (usually) states possible at any point, e.g., a transistor that can either be in saturation or be nonconducting. We usually choose to talk about voltages rather than currents, calling a level HIGH or LOW. The two states can represent any of a variety of “bits” (binary digits) of information, such as the following:

- one bit of a number
- whether a switch is opened or closed
- whether a signal is present or absent
- whether some analog level is above or below some preset limit
- whether or not some event has happened yet
- whether or not some action should be taken
- etc.

HIGH and LOW

The HIGH and LOW states represent the TRUE and FALSE states of Boolean logic, in some predefined way. If at some point HIGH represents TRUE, that signal line is called “positive true,” and vice versa. This can be confusing at first. Figure 8.1 shows an example. SWITCH CLOSED is true when the output is LOW; that’s a negative-true signal (“LOW-true” might be a better label, since no negative voltages are involved), and you might label the lead as shown (a bar over a symbol means NOT; that line is HIGH when the switch is *not* closed). Just remember that the presence or absence of the negation bar over the label tells whether the wire is LOW or HIGH when the stated condition (SWITCH CLOSED) is true.

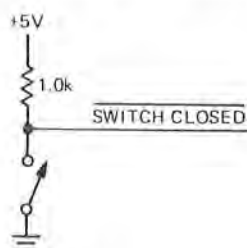


Figure 8.1

A digital circuit “knows” what a signal represents by where it comes from, just as an analog circuit might “know” what the output of some op-amp represents. However, added flexibility is possible in digital circuits; sometimes the same signal lines are used to carry different kinds of information, or even to send it in different directions, at different times. In order to do this “multiplexing,” additional information must also be sent (address bits, or status bits). You will see many examples of this very useful ability later. For now, imagine that any given circuit is wired up to perform a predetermined function and that it knows what that

function is, where its inputs are coming from, and where the outputs are going.

To lend a bit of confusion to a basically simple situation, we introduce 1 and 0. These symbols are used in Boolean logic to mean TRUE and FALSE, respectively, and are sometimes used in electronics in exactly that way. Unfortunately, they are also used in another way, in which 1 = HIGH and 0 = LOW! In this book we will try to avoid any ambiguity by using the word HIGH (or the symbol H) and the word LOW (or the symbol L) to represent logic states, a method that is in wide use in the electronics industry. We will use 1 and 0 only in situations where there can be no ambiguity.

Voltage range of HIGH and LOW

In digital circuitry, the voltage levels corresponding to HIGH and LOW are allowed to fall in some range. For example, with high-speed CMOS ("HC") logic, *input* voltages within about 1.5 volts of ground are interpreted as LOW, while voltages within 1.5 volts of the +5 volt supply are interpreted as HIGH. In fact, typical LOW- and HIGH-state *output* voltages are usually within a tenth of a volt of 0 and +5 volts, respectively (the output is a saturated MOS transistor to one of the rails; see Fig. 8.17). This allows for manufacturing spread, variations of the circuits with temperature, loading, supply voltage, etc., and the presence of "noise," the miscellaneous garbage that gets added to the signal in its journey through the circuit (from capacitive coupling, external interference, etc.). The circuit receiving the signal decides if it is HIGH or LOW and acts accordingly. As long as noise does not change 1's to 0's, or vice versa, all is well, and any noise is eliminated at each stage, since "clean" 0's and 1's are regenerated. In that sense digital electronics is noiseless and perfect.

The term *noise immunity* is used to describe the maximum noise level that can be added to logic levels (in the worst case) while still maintaining error-free operation. For instance, TTL has 0.4 volt noise immunity, since a TTL *input* is guaranteed to interpret anything less than +0.8 volt as LOW and anything greater than +2.0 volts as HIGH, whereas the worst-case *output* levels are +0.4 volt and +2.4 volts, respectively (see the accompanying box on logic levels). In practice, noise immunity is considerably better than this figure, with typical LOW and HIGH voltages of +0.2 and +3.4 volts and an input decision threshold near +1.3 volts. But always remember that if you are doing good circuit design, you use worst-case values. It is worth keeping in mind that different logic families have different amounts of noise immunity. CMOS has greater voltage noise immunity than TTL, whereas the speedy ECL family has less. Of course, susceptibility to noise in a digital system depends also on the amplitude of noise that is present, which in turn depends on factors such as output-stage stiffness, inductance in the ground leads, the existence of long "bus" lines, and output slew rates during logic transitions (which produces transient currents, and therefore voltage spikes on the ground line, due to capacitive loading). We will worry about some of these problems in Sections 9.11–9.13.

8.03 Number codes

Most of the conditions we listed earlier that can be represented by a digital level are self-explanatory. How a digital level can represent part of a number is a more involved, and very interesting, question.

A decimal (base-10) number is simply a string of integers that are understood to multiply successive powers of 10, the individual products then being added together.

For instance,

$$137.06 = 1 \times 10^2 + 3 \times 10^1 + 7 \times 10^0 \\ + 0 \times 10^{-1} + 6 \times 10^{-2}$$

Ten symbols (0 through 9) are needed, and the power of 10 each multiplies is determined by its position relative to the decimal point. If we want to represent a number using two symbols only (0 and 1), we use the *binary*, or base-2, number system. Each 1 or 0 then multiplies a successive power of 2. For instance,

$$1101_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ = 13_{10}$$

The individual 1's and 0's are called "bits" (binary digits). The subscript (always given in base 10) tells what number system we are using, and often it is essential in order to avoid confusion, since the symbols all look the same.

We convert a number from binary to decimal by the method just described. To convert the other way, we keep dividing the number by 2, and write down the remainders. To convert 13_{10} to binary,

$$\begin{aligned} 13/2 &= 6 \text{ remainder } 1 \\ 6/2 &= 3 \text{ remainder } 0 \\ 3/2 &= 1 \text{ remainder } 1 \\ 1/2 &= 0 \text{ remainder } 1 \end{aligned}$$

from which $13_{10} = 1101_2$. Note that the answer comes out in the order LSB (least significant bit) to MSB (most significant bit).

Hexadecimal ("hex") representation

The binary number representation is the natural choice for two-state systems (although it is not the only way; you'll see some others soon). Since the numbers tend to get rather long, it is common to write them in hexadecimal (base-16) representation: Each position represents successive powers of 16, with each hex symbol

having a value from 0 to 15. Since you want a single symbol for each hex position, the symbols A–F are assigned to the values 10–15. To write a binary number in hexadecimal, just group it in 4-bit groups, beginning with the LSB, and write the hexadecimal equivalent of each group:

$$707_{10} = 1011000011_2 (= 10\ 1100\ 0011_2) \\ = 2C3_{16}$$

Hexadecimal representation is well suited to the popular "byte" (8-bit) organization of computers, which are most often organized as 16-bit or 32-bit computer "words"; a word is then 2 or 4 bytes. An alphanumeric character (letter, number, or symbol) is one byte. So in hexadecimal, each byte is 2 hex digits, a 16-bit word is 4 hex digits, etc.

For example, in the widely used ASCII representation (more on that in Section 10.19), lower-case "a" is ASCII value 01100001 (61 hex, which we will write as 61_H), "b" is 62_H , etc. Thus the word "nerd" could be stored in a pair of 16-bit words, whose hex values are $6E65_H$ and 7264_H . As another example, the memory locations in a computer with 65,536 ("64K") bytes of memory can be identified by a 2-byte address, since $2^{16} = 65,536$; the lowest address is 0000_H , the highest address is $FFFF_H$, the second half of memory begins at 8000_H , and the fourth quarter of memory begins at $C000_H$.

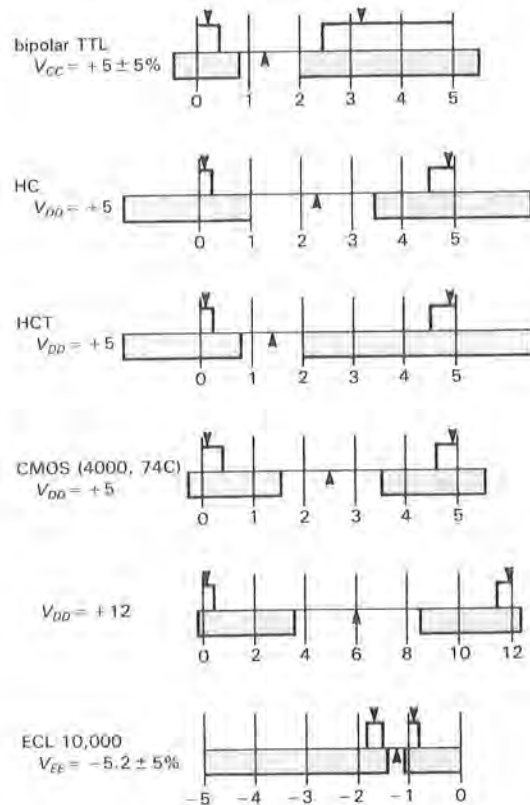
You occasionally see "octal" (base-8) notation, a relic of an earlier era when computers used 12-bit and 36-bit words, with 6-bit alphanumeric representation. Although octal has the comfortable feature of using only familiar symbols (0–7), it is extremely awkward when applied to byte-organized words. Exercise 8.1 shows you why.

EXERCISE 8.1

Begin by writing down the octal representation for ASCII "a" and "b" using the hex values

LOGIC LEVELS

The diagram shows the ranges of voltages that correspond to the two logic states (HIGH and LOW) for the most popular families of digital logic. For each logic family it is necessary to specify legal values of both output and input voltages corresponding to the two states HIGH and LOW. The shaded areas above the line show the specified range of output voltages that a logic LOW or HIGH is guaranteed to fall within, with the pair of arrows indicating typical output values (LOW, HIGH) encountered in practice. The shaded areas below the line show the range of input voltages guaranteed to be interpreted as LOW or HIGH, with the arrow indicating the typical *logic threshold* voltage, i.e., the dividing line between LOW and HIGH. In all cases a logic HIGH is more positive than a logic LOW.



The meanings of "minimum," "typical," and "maximum," in electronic specifications are worth a few words of explanation. Most simply, the manufacturer guarantees that the components will fall in the range minimum–maximum, with many close to "typical." What this means is that typical specifications are what you use when designing circuits; however, those circuits must work properly over the whole range of specifications from minimum to maximum (the extremes of manufacturing variability). In particular, a well-designed circuit must function under the worst possible combination of minimum and maximum values. This is known as *worst-case design*, and it is essential for any instrument produced from off-the-shelf (i.e., not specially selected) components.

given earlier. Then write down the octal representation of the 16-bit word formed by putting the two bytes for "ab" together. Why are the individual identities of the characters lost? What is "ba" in octal? Now do the same things, but using hexadecimal notation.

BCD

Another way to represent a number is to encode each decimal digit into binary. This is called BCD (binary-coded decimal), and it requires a 4-bit group for each digit. For instance,

$$137_{10} = 0001\,0011\,0111 \quad (\text{BCD})$$

Note that BCD representation is *not* the same as binary representation, which in this case would be $137_{10} = 10001001_2$. You can think of the bit positions (starting from the right) as representing 1, 2, 4, 8, 10, 20, 40, 80, 100, 200, 400, 800, etc. It is clear that BCD is wasteful of bits, since each 4-bit group could represent numbers 0 through 15, but in BCD never represents numbers greater than 9. However, BCD is ideal if you want to display a number in decimal, since all you do is convert each BCD character to the appropriate decimal number and display it. (There are many devices that do exactly this, e.g., a "BCD decoder, driver, and display," which is a little IC with a transparent top. You apply logic levels for your BCD character, and it lights up with the digit.) For this reason, BCD is commonly used for input and output of numeric information. Unfortunately, the conversion between pure binary and BCD is complicated, since *each* decimal digit depends on the state of almost every binary bit, and vice versa. Nevertheless, binary arithmetic is so efficient that most computers convert all input data to binary, converting back only when data need to be output. Think how much effort and bother would have been saved if *Homo sapiens* had evolved with 8 (or 16) fingers!

EXERCISE 8.2

Convert to decimal: (a) 1110101.0110_2 , (b) $11.01010101\dots_2$, (c) $2A_H$. Convert to binary: (a) 1023_{10} , (b) 1023_H . Convert to hexadecimal: (a) 1023_{10} , (b) 101110101101_2 , (c) 61453_{10} .

Signed numbers

Sign magnitude representation. Sooner or later it becomes necessary to represent negative numbers in binary, particularly in devices where some computation is done. The simplest method is to devote one bit (the MSB, say) to the sign, with the remaining bits representing the magnitude of the number. This is called "sign magnitude representation," and it corresponds to the way signed numbers are ordinarily written (see Table 8.1). It is used when numbers are displayed, as well as in some A/D conversion schemes. In general, it is not the best method for representing signed numbers, particularly where some computation is done, for several reasons: Computation is awkward; subtraction is different from addition (i.e., addition doesn't "work" for signed numbers). Also, there can be two zeros (+0 and -0), so you have to be careful to use only one of them.

Offset binary representation. A second method for representing signed numbers is "offset binary," in which you subtract half the largest possible number to get the value represented (Table 8.1). This has the advantage that the number sequence from the most negative to the most positive is a simple binary progression, which makes it a natural for binary "counters." The MSB still carries the sign information, and zero appears only once. Offset binary is popular in A/D and D/A conversions, but it is still awkward for computation.

2's complement representation. The method most widely used for integer computation is called "2's complement."

TABLE 8.1. 4-BIT SIGNED INTEGERS IN THREE SYSTEMS OF REPRESENTATION

Integer	Sign-magnitude	Offset binary	2's comp
+7	0111	1111	0111
+6	0110	1110	0110
+5	0101	1101	0101
+4	0100	1100	0100
+3	0011	1011	0011
+2	0010	1010	0010
+1	0001	1001	0001
0	0000	1000	0000
-1	1001	0111	1111
-2	1010	0110	1110
-3	1011	0101	1101
-4	1100	0100	1100
-5	1101	0011	1011
-6	1110	0010	1010
-7	1111	0001	1001
-8	—	0000	1000
(-0)	1000	—	—

In this system, positive numbers are represented as simple unsigned binary. The system is rigged up so that a negative number is then simply represented as the binary number you add to a positive number of the same magnitude to get zero. To form a negative number, first complement each of the bits of the positive number (i.e., write 1 for 0, and vice versa; this is called the "1's complement"), then add 1 (that's the "2's complement"). As you can see from Table 8.1, 2's complement numbers are related to offset binary numbers by having the MSB complemented. As with the other signed number representations, the MSB carries the sign information. There's only one zero, conveniently represented by all bits 0 ("clearing" a counter or register sets its value to zero).

Arithmetic in 2's complement

Arithmetic is simple in 2's complement. To add two numbers, just add bitwise (with carry), like this:

$$5 + (-2) : \begin{array}{r} 0101 \quad (+5) \\ 1110 \quad (-2) \\ \hline 0011 \quad (+3) \end{array}$$

To subtract B from A , take the 2's complement of B and add (i.e., add the negative):

$$2 - 5 : \begin{array}{r} 0010 \quad (+2) \\ 1011 \quad (-5) \end{array}$$

$$\begin{array}{r} (+5 = 0101 : 1's \\ \text{comp} = 1010, \text{ so} \\ 2's \text{ comp} = 1011) \end{array}$$

$$\begin{array}{r} 1101 \quad (-3) \end{array}$$

Multiplication also "works right" in 2's complement representation. Try the following exercise.

EXERCISE 8.3

Multiply +2 by -3 in 3-bit 2's complement binary arithmetic. Hint: The answer is -6.

EXERCISE 8.4

Show that the 2's complement of -5 is +5.

Because the 2's complement system is natural for computation, it is universally used for integer arithmetic in computers (note, however, that "floating point" numbers are usually represented in a form of "sign magnitude," namely sign-exponent-mantissa).

□ Gray code

The following code is used for mechanical shaft-angle encoders, among other things. It is called a Gray code, and it has the property that only one bit changes in going from one state to the next. This prevents errors, since there is no way of guaranteeing that all bits will change simultaneously at the boundary between two encoded values. If straight binary were used, it would be possible to generate an output of 15 in going from 7 to 8, for instance. Here is a simple rule for

generating Gray-code states: Begin with a state of all zeros. To get to the next state, always change the single least significant bit that brings you to a new state.

0000
0001
0011
0010
0110
0111
0101
0100
1100
1101
1111
1110
1010
1011
1001
1000

Gray codes can be generated with any number of bits. They find use also in “parallel encoding,” a technique of high-speed A/D conversion that you will see later. We will talk about translation between Gray-code and binary-code representations in the next section.

8.04 Gates and truth tables

Combinational versus sequential logic

In digital electronics the name of the game is generating digital outputs from digital inputs. For instance, an *adder* might take two 16-bit numbers as inputs and generate a 16-bit (plus carry) sum. Or you might build a circuit to multiply two numbers. These are the kinds of operations a computer’s processing unit should be able to do. Another task might be to compare two numbers to see which is larger or to compare a set of inputs with the desired input to make sure that “all systems are go.” Or you might want to attach a “parity bit” to a number to make the total number of 1’s even, say, before transmission over a data link; then the parity could be

checked on receipt as a simple check of correct transmission. Another typical task is to take some numbers expressed in binary and display, print, or punch them as decimal characters. All of these are tasks in which the output or outputs are predetermined functions of the input or inputs. As a class, they are known as “combinational” tasks. They can all be performed with devices called *gates*, which perform the operations of Boolean algebra applied to two-state (binary) systems.

There is a second class of problems that cannot be solved by forming a combinational function of the inputs alone, but require knowledge of past inputs as well. Their solution requires the use of “sequential” networks. Typical tasks of this type might be converting a string of bits in serial form (one after another) into a parallel set of bits, or keeping count of the number of 1’s in a sequence, or recognizing a certain pattern in a sequence, or giving one output pulse for each four input pulses. All these tasks require digital memory of some sort. The basic device here is the “flip-flop” (the fancy name is “bistable multivibrator”).

We will begin with gates and combinational logic, since they’re basic to everything. Digital life will become more interesting when we get to sequential devices, but there will be no lack of fun and games with gates alone.

OR gate

The output of an OR gate is HIGH if either input (or both) is HIGH. This can be expressed in a “truth table,” as shown in Figure 8.2. The gate illustrated is a 2-input OR gate. In general, gates can have any number of inputs, but the standard packages usually contain four 2-input gates, three 3-input gates, or two 4-input gates. For instance, a 4-input OR gate will have a HIGH output if any one input (or more) is HIGH.

The Boolean symbol for OR is $+$. "A OR B" is written $A + B$.

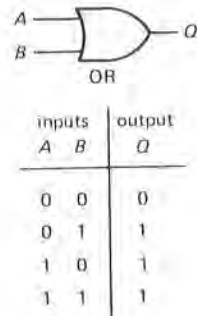


Figure 8.2

AND gate

The output of an AND gate is HIGH only if both inputs are HIGH. The logic symbol and truth table are as shown in Figure 8.3. As with OR gates, AND gates are available with 3 or 4 (sometimes more) inputs. For instance, an 8-input AND gate will have a HIGH output only if *all* inputs are HIGH.

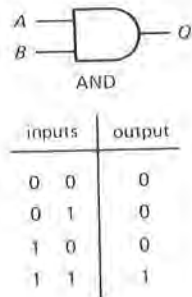


Figure 8.3

The Boolean symbol for AND is a dot ($\cdot$); this can be omitted, and usually is. "A AND B" is written $A \cdot B$, or simply AB .

Inverter (the NOT function)

Frequently we need the complement of a logic level. That is the function of

an inverter, a "gate" with only one input (Fig. 8.4).

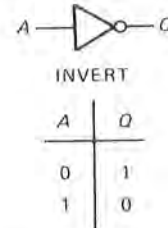


Figure 8.4

The Boolean symbol for NOT is a bar over the symbol, or sometimes a prime symbol. "NOT A" is written $\bar{A}$, or A' . For the convenience of typesetters, the symbols $/$, $*$, $-$, and $'$ are often used, in place of the overbar, to indicate NOT; thus, "NOT A" might be written as any of the following: A' , $-A$, $*A$, $/A$, A^* , $A/$. A given document will usually pick one of these alternatives and stick with it throughout. We have chosen the form A' for this book.

NAND and NOR

The INVERT function can be combined with gates, forming NAND and NOR (Fig. 8.5). These are actually more popular than AND and OR, as you will see shortly.

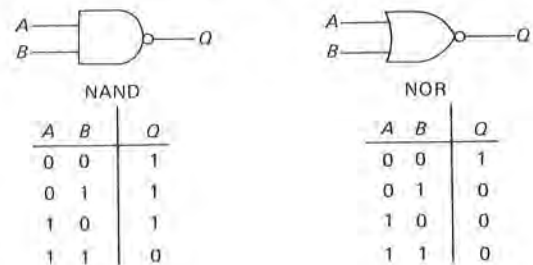


Figure 8.5

Exclusive-OR

Exclusive-OR is an interesting function, although less fundamental than AND and

OR (Fig. 8.6). The output of an exclusive-OR gate is HIGH if one or the other (but not both) input is HIGH (it never has more than two inputs). Another way to say it is that the output is HIGH if the inputs are different. The exclusive-OR gate is identical with modulo-2 addition of two bits.

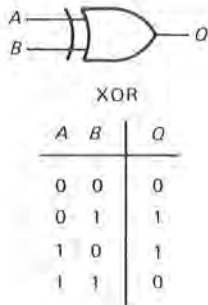


Figure 8.6

EXERCISE 8.5

Show how to use the exclusive-OR gate as an "optional inverter," i.e., it inverts an input signal or buffers it without inversion, depending on the level at a control input.

EXERCISE 8.6

Verify that the circuits in Figure 8.7 convert binary code to Gray code, and vice versa.

□ 8.05 Discrete circuits for gates

Before going on to discuss gate applications, let's see how to make gates from discrete components. Figure 8.8 shows a diode AND gate. If either input is held LOW, the output is LOW. The output can go HIGH only when both inputs go HIGH. This circuit has many disadvantages. In particular: (a) Its LOW output is a diode drop above the signal holding the input LOW. Obviously you couldn't use very many of these in a row! (b) There is no "fanout" (the ability of one output to

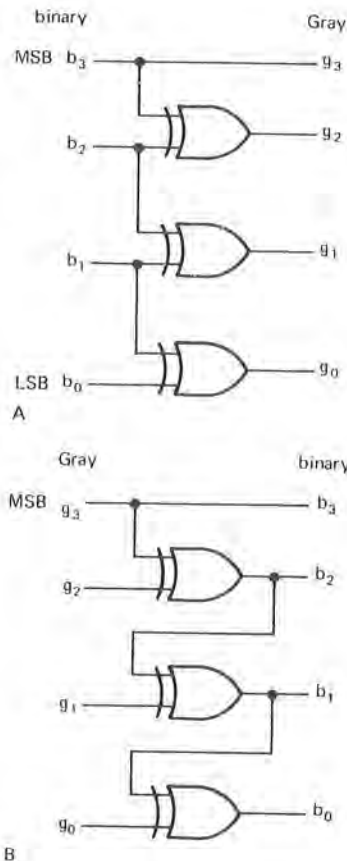


Figure 8.7. Parallel code converters: binary to Gray and Gray to binary.

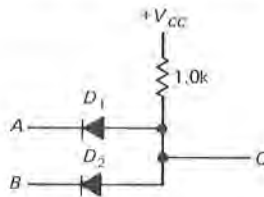


Figure 8.8

drive several inputs), since any load at the output is seen by the signal at the input. (c) It is slow, because of resistive pull-up. As a general rule, you cannot do as well with logic constructed from discrete components as with IC logic. Part of the superiority of IC logic lies in the use of special techniques (e.g., ion

implantation) to achieve excellent performance.

Figure 8.9 shows the simplest form of transistor NOR gate. This circuit was used in the family of logic known as RTL (resistor-transistor logic), which was popular in the 1960s because of its low price, but is now obsolete. A HIGH at either input (or both) turns on at least one transistor, pulling the output LOW. Since this gate is intrinsically inverting, you would have to add an inverter, as shown, to make an OR gate.

Although the discrete gate circuits just illustrated are simple to understand, you wouldn't use them in practice because of their disadvantages. In fact, except in rare circumstances you would never construct gates (or any other logic) from discrete components, since a full range of excellent logic is available as inexpensive and compact integrated circuits, as we will see shortly. Currently the most popular IC logic circuits are built with complementary MOSFETs ("CMOS"). Look back at Figure 3.59 to remind yourself how you would make a CMOS NAND gate.

8.06 Gate circuit example

Let's work out a circuit to perform the logic we gave as an example in Chapters

1 and 2: the task to sound a buzzer if either car door is open and the driver is seated. The answer is obvious if you restate the problem as "output HIGH if either the left door OR the right door is open, AND driver is seated," i.e., $Q = (L + R)S$. Figure 8.10 shows it with gates. The output of the OR gate is HIGH if one OR the other door (or both) is open. If that is so, AND the driver is seated, Q goes HIGH. With an additional transistor, this could be made to sound a buzzer or close a relay.

In practice, the switches generating the inputs will probably close a circuit to ground, to save extra wiring (there are additional reasons, particularly in the case of the popular TTL logic, and we will get to them shortly). This means that the inputs will go LOW when a door is opened, for example. In other words, we have "negative-true" inputs. Let's rework the example with this in mind, calling the inputs L' , R' , and S' .

First, we need to know if either door input (L' , R') is LOW; i.e., we must distinguish the state "both inputs HIGH" from all others. That's an AND gate. So we make L' and R' the inputs to an AND gate. The output will be LOW if either input is LOW; call that EITHER'. Now we need to know when EITHER' is LOW and S'

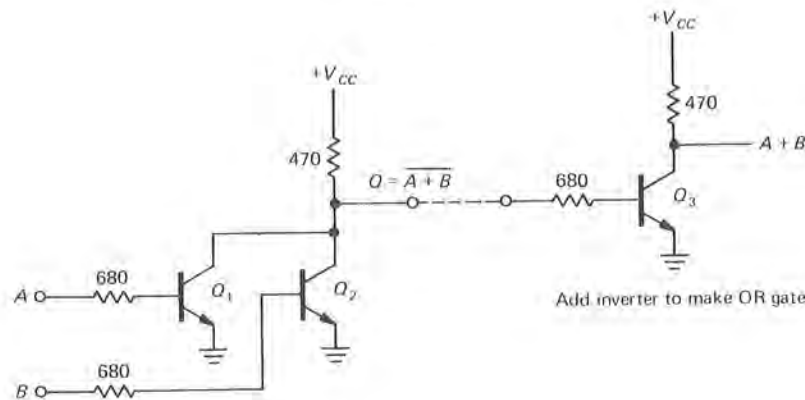


Figure 8.9

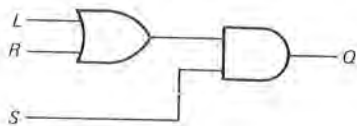


Figure 8.10

is LOW; i.e., we must distinguish the state “both inputs LOW” from all others. That’s an OR gate. Figure 8.11 shows the circuit. We have used a NOR gate, instead of an OR gate, to get the same output as earlier: Q HIGH when the desired condition is present. Something strange seems to be going on here, though. We have used AND instead of OR (and vice versa), as compared with the earlier circuit. Section 8.07 should clarify the matter. First, consider the following exercise.

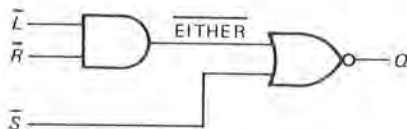


Figure 8.11

EXERCISE 8.7

What do the circuits shown in Figure 8.12 do?

Gate interchangeability

When designing digital circuits, keep in mind that it is possible to form one kind of gate from another. For example, if you need an AND gate, and you have half of a 7400 available (quad 2-input NAND), you can substitute as shown in Figure 8.13. The second NAND functions as an inverter, making AND. The following exercises should help you explore this idea.

EXERCISE 8.8

Using 2-input gates, show how to make (a) INVERT from NOR, (b) OR from NORs, and (c) OR from NANDs.

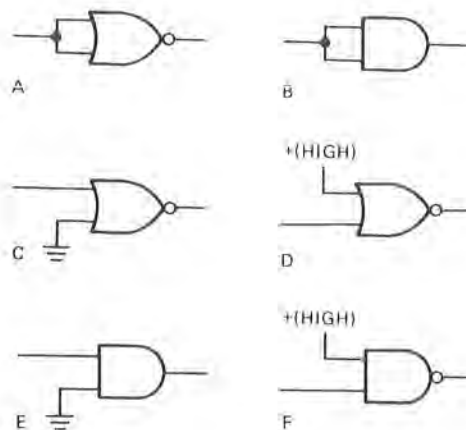


Figure 8.12

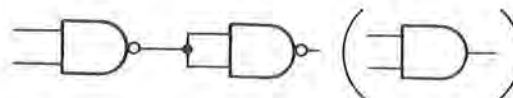


Figure 8.13

EXERCISE 8.9

Show how to make (a) a 3-input AND from 2-input ANDs, (b) a 3-input OR from 2-input ORs, (c) 3-input NOR from 2-input NORs, and (d) a 3-input AND from 2-input NANDs.

In general, multiple use of one kind of inverting gate (e.g., NAND) is enough to make any combinational function. However, this isn’t true for a noninverting gate, since there’s no way to make INVERT. This probably accounts for the greater popularity of NAND and NOR in logic design.

8.07 Assertion-level logic notation

An AND gate has a HIGH output if both inputs are HIGH. So, if HIGH means “true,” you get a true output only if all inputs are true. In other words, with positive-true logic, an AND gate performs the AND function. The same holds for OR.

What happens if LOW means “true,” as in the last example? An AND gate gives a LOW if either input is true (LOW): It’s an OR function! Similarly an OR gate gives a LOW only if both inputs are true (LOW). It’s an AND function! Very confusing.

There are two ways to handle this problem. The first way is to think through any digital design problem as we did earlier, choosing the kind of gate that gives the needed output. For instance, if you need to know if any of three inputs is LOW, use a 3-input NAND gate. This method is still used by some misguided designers. When designing this way, you would draw a NAND gate, even though the gate is performing a NOR function on the (negative-true) inputs. You would probably label the inputs as in Figure 8.14. In this example, $\overline{\text{CLEAR}}'$, $\overline{\text{MR}}'$ (master reset), and $\overline{\text{RESET}}'$ might be negative-true levels coming from various places in a circuit. The output, CLR, is positive-true and will go to the devices that are to be cleared if *any* of the reset signals goes LOW (true).

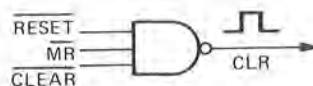


Figure 8.14

The second way to handle the problem of negative-true signals is to use “assertion-level logic.” If a gate performs an OR function on negative-true inputs, draw it that way, as in Figure 8.15. The 3-input OR gate with negated inputs is functionally identical with the preceding 3-input NAND. That equivalence turns out to be an important logical identity, as stated in DeMorgan’s theorem, and we will spell out a number of such useful identities shortly. For now, it is enough to know that you can change AND to OR (and

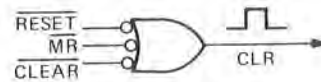


Figure 8.15

vice versa) if you negate the output and all inputs (see Table 8.2). Assertion-level logic looks forbidding at first, because of the proliferation of funny-looking gates. It is better, though, because the logical functions of the gates in the circuit stand out clearly. You’ll find it “friendly” after you’ve used it for a while, and you won’t want to use anything else. Let’s rework the car door example again with assertion-level logic (Fig. 8.16). The gate on the left determines if L or R is true, i.e., LOW, giving a negative-true output. The second gate gives a HIGH output if both $(L + R)$ and S are true, i.e., LOW. From DeMorgan’s theorem (after a while you won’t even need that, you’ll recognize these gates as equivalent) the first gate is AND, the second is NOR, just as in the circuit drawn earlier. Two important points:

1. Negative-true doesn’t mean that the logic levels are negative *polarity*. It means that the lower of the two states (LOW) stands for TRUE.
2. The symbol used to draw the gate itself assumes positive-true logic. A NAND gate used as an OR for negative-true signals can be drawn as a NAND or, using assertion-level logic, as an OR with negation symbols (little circles) at the inputs. In the latter case you think of the circles as indicating inversion of the input signals, followed by an OR gate operating on positive-true logic as originally defined.

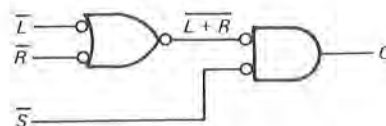


Figure 8.16

TABLE 8.2 COMMON GATES IN THE TTL AND CMOS FAMILIES

Name	Expression	Symbol	Negative true symbol	Type	No. per chip	Part number									
						4000B CMOS	74xx	→	ALS	AS	F	LS	C	AC(T)	HC(T)
AND	AB			2-input	4	4081	7408		✓	✓	✓	✓	✓	✓	✓
				3-input	3	4073	7411		✓	✓	✓	✓	✓	✓	✓
				4-input	2	4082	7421		✓	✓	✓	✓	✓	✓	✓
NAND	$\overline{AB}$			2-input	4	4011	7400		✓	✓	✓	✓	✓	✓	✓
				3-input	3	4023	7410		✓	✓	✓	✓	✓	✓	✓
				4-input	2	4012	7420		✓	✓	✓	✓	✓	✓	✓
				8-input	1	4068	7430		✓	✓	✓	✓	✓	✓	✓
				13-input	1	—	74133		✓	✓	✓	✓	✓	✓	✓
OR	$A + B$			2-input	4	4071	7432		✓	✓	✓	✓	✓	✓	✓
				3-input	3	4075	—								
				4-input	2	4072	74802			✓					✓
NOR	$\overline{A + B}$			2-input	4	4001	7402		✓	✓	✓	✓	✓	✓	✓
				3-input	3	4025	7427		✓	✓	✓	✓	✓	✓	✓
				4-input	2	4002	7425								✓
				5-input	2	—	74260				✓				✓
				8-input	1	4078	—								✓
INVERT	$\overline{A}$				6	4069/4049	7404		✓	✓	✓	✓	✓	✓	✓
					8	—	74240		✓	✓	✓	✓	✓	✓	✓
BUFFER	A				6	4503/4050	74365		✓	✓	✓	✓	✓	✓	✓
					8	—	74241/244 (-541/-544)		✓		✓	✓	✓	✓	✓
XOR	$A \oplus B$			2-input	4	4070	7486/386 (-135)		✓		✓	✓	✓	✓	✓
XNOR	$\overline{A \oplus B}$			2-input	4	4077	74266 (-135)					✓			✓
AOI				2-2-input	2	4085	7450/51				✓	✓			✓
				2-2-2-2-input	1	4086	7453/54					✓			✓

Postscript: Logical AND and OR shouldn't be confused with the *legal* equivalents. The weighty legal tome known as *Words and Phrases* has over 40 pages of situations in which AND can be construed as OR. For example: "OR will be construed AND, and AND will be construed OR, as the necessities of the case may require. . . ." This isn't the same as DeMorgan's theorem!

TTL AND CMOS

TTL (transistor-transistor logic) and CMOS (complementary MOS) are the two most popular logic families in current use, with at least 10 manufacturers of integrated

circuits offering an enormous variety of functions in both families. These families should satisfy your needs for all digital design, with the exceptions of some large-scale integration (LSI), which uses either CMOS or NMOS logic, and ultra-high-speed logic, where GaAs devices and emitter-coupled logic (ECL) reign supreme. Throughout the rest of the book we will rely heavily on these families.

8.08 Catalog of common gates

Table 8.2 shows the common gates you can get in the TTL and CMOS families of digital logic. Each gate is drawn in its normal (positive-true) incarnation, and

also the way it looks for negative-true logic. The last entry in the table is an AND-OR-INVERT gate, sometimes abbreviated AOI.

A word of explanation: Digital logic is available in 10 popular subfamilies (CMOS: 4000B, 74C, 74HC, 74HCT, 74AC, 74ACT; and TTL: 74LS, 74ALS, 74AS, 74F), all offering the same functions and with a pretty good degree of compatibility between them. The differences have to do with speed, power dissipation, output drive capability, and logic levels (see Sections 8.09 and 9.02). The best type for most applications is currently “high-speed CMOS,” specified by adding the letters HC after the digits 74, e.g., 74HC00. Where compatibility with existing bipolar TTL is required, however, you should use the HCT (or possibly LS) subfamily. For simplicity we will routinely omit such letters (and the 74- prefix) in this book, indicating digital IC types with an apostrophe, e.g., '00 for a 2-input NAND. Note that the original TTL (“7400 family” – no letters after the “74”) is obsolete. We’ll describe the interesting history of these families in Section 9.01.

8.09 IC gate circuits

Although a NAND gate, for instance, performs identical logic operations in the various TTL and CMOS versions, the logic levels and other characteristics (speed, power, input current, etc.) are quite different. In general, you have to be careful when mixing logic family types. To understand the differences, look at the schematics of a NAND gate in Figure 8.17.

The CMOS gate is constructed from enhancement-mode MOSFETs of both polarities, connected as switches rather than followers. An ON FET looks like a low resistance to whichever supply rail it is connected. Both inputs must be HIGH to turn on the series pair Q_3Q_4 and to turn off both of the pull-up transistors Q_1Q_2 . That produces a LOW at the output, i.e., it

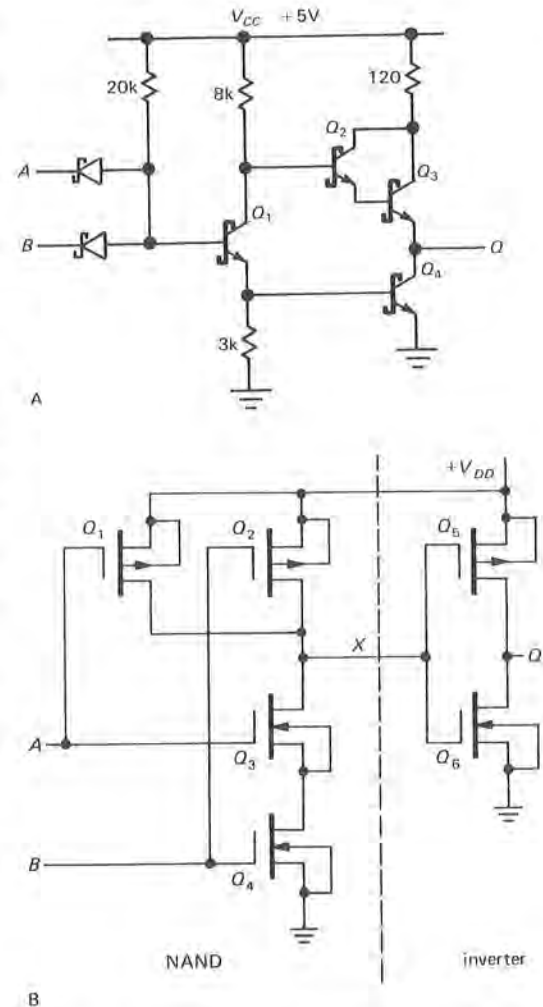


Figure 8.17. A. LS TTL NAND gate.
B. CMOS AND gate.

is a NAND gate. Q_5 and Q_6 constitute the standard CMOS inverter, to generate an AND gate. From this example it should be evident how to generalize to AND, NAND, OR, and NOR with any number of inputs.

EXERCISE 8.10

Draw the circuit of a 3-input CMOS OR gate.

The bipolar LS (low-power Schottky) TTL NAND basically consists of the diode-resistor logic of Figure 8.8 driving a transistor inverter followed by a push-pull

output. If both inputs are HIGH, the 20k resistor holds Q_1 on, thus producing a LOW output by saturating Q_4 and shutting off Darlington Q_2Q_3 . If at least one input is LOW, Q_1 is held off, thus producing a HIGH output by follower action of Q_2Q_3 combined with Q_4 being held off. Schottky diodes and Schottky-clamped transistors are used throughout for enhanced speed.

Note that both CMOS and bipolar TTL gates have an output circuit with “active pullup” to the positive supply rail, unlike our discrete gate examples.

8.10 TTL and CMOS characteristics

Let's compare family characteristics:

Supply voltage. The bipolar TTL families require +5 volts, $\pm 5\%$, whereas the CMOS families have a wider range: +2 to +6 volts for HC and AC, +3 to +15 volts for 4000B and 74C. The HCT and ACT CMOS families, designed for compatibility with bipolar TTL (see below), require +5 volts.

Input. A TTL input held in the LOW state sources current into whatever drives it (0.25mA typ for LS), so to pull it LOW you must sink current. Since the TTL output circuit (a saturated *npn* transition) is good at sinking current, this presents no problem when TTL logic is wired together, but you must keep it in mind when driving TTL with other circuitry. By contrast, CMOS has no input current.

The TTL input logic threshold is about two diode drops above ground (about 1.3V), whereas most CMOS families have their threshold nominally at half the supply voltage (though with considerable spread, typically 1/3 to 2/3 the supply voltage). The HCT and ACT CMOS families are designed with a low threshold similar to bipolar TTL for compatibility, since a bipolar TTL output does not swing all the way to +5 volts (see below).

CMOS inputs are susceptible to damage from static electricity during handling. In both families, unused inputs should be tied HIGH or LOW, as necessary (more on this later).

Output. The TTL output stage is a saturated transistor to ground in the LOW state, and a (Darlington) follower in the HIGH state (two diode drops below V_+). For all CMOS families (including HCT and ACT) the output is a turned-on MOSFET, either to ground or to V_+ ; i.e., rail-to-rail output swings. In general, the faster families (F, AS, AC, ACT) have greater output drive capability than the slower families (LS, 4000B, 74C, HC, HCT).

Speed and power. The bipolar TTL families consume considerable quiescent current, more for the faster families (AS and F); the corresponding speeds go from about 25MHz (for LS) to about 100MHz (for AS and F). All CMOS families consume zero quiescent current. However, their power consumption rises linearly with increasing frequency (switching capacitive loads requires current), and CMOS operated near its upper frequency limit often dissipates as much power as the equivalent bipolar TTL family (Fig. 8.18). The speed range of CMOS goes from about 2MHz (for 4000B/74C at 5V) to about 100MHz (for AC/ACT).

In general, the nice characteristics of CMOS (zero quiescent current, rail-to-rail output swings, good noise immunity) make it the logic of choice, and we recommend the HC family for most new designs. However, for greater speed, use AC; for wide supply range where high speed is not needed, use 74C or 4000B; use HCT (or perhaps LS) for compatibility with bipolar TTL outputs, unless you need the speed of ACT (or AS or F). In some

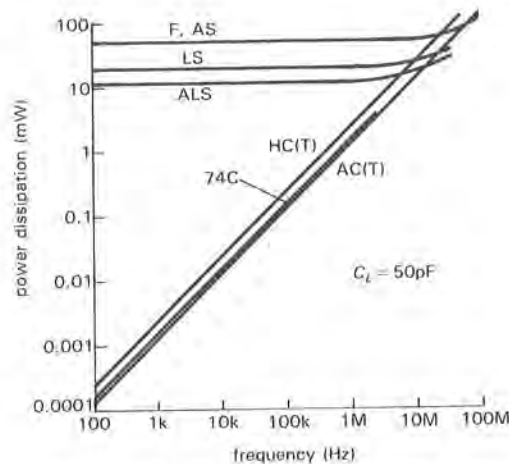


Figure 8.18. Gate power dissipation versus frequency.

high-density applications (memory, microprocessor), NMOS devices are preferred, in spite of their relatively high power dissipation. And for the highest-speed applications (above 100MHz) you are forced to use either ECL, which goes up to about 500MHz, or logic based on GaAs, which is usable to about 4GHz. See Section 14.15 and Table 9.1 for further discussion of CMOS logic families.

Within any one logic family, outputs are designed to drive other inputs easily, so you don't often have to worry about thresholds, input current, etc. For instance, with TTL or CMOS, any output can drive at least 10 other inputs (the official term for this is *fanout*: TTL has a fanout of 10), so you don't have to do anything special to ensure compatibility. In the next chapter we will go into the issue of interfacing between logic families and between logic circuits and the outside world.

8.11 Three-state and open-collector devices

The TTL and CMOS gates we have just discussed have push-pull output circuits: The output is held either HIGH or LOW by an ON transistor or MOSFET. Nearly

all digital logic uses this sort of circuit (called active pullup; in TTL it's also called a totem-pole output) because it provides low output impedance in both states, giving faster switching time and better noise immunity, as compared with an alternative such as a single transistor with passive collector pullup resistor. In the case of CMOS, it also results in lower power dissipation.

However, there are a few situations for which active pullup output is unsuitable. As an example, imagine a computer system in which several functional units have to exchange data. The central processor (CPU), memory, and various peripherals all need to be able to send and receive 16-bit words. It would be awkward (to put it mildly) to have separate 16-wire cables connecting each device to all others. The solution is the so-called *data bus*, a single set of 16 wires accessible to all devices. It's like a telephone party line: Only one device at a time may "talk" (assert data), but all may "listen" (receive data). With a bus system there must be agreement as to who may talk, and words like "bus arbitrator," "bus master," and "control bus" pop up.

You can't use gates (or any other devices) with active pullup outputs to drive a bus, since you couldn't disconnect your output from the shared data lines (you're holding it either HIGH or LOW at all times). What's needed is a gate whose output can be "open." Such devices are available, and they come in two varieties, "three-state devices" and "open-collector devices."

Three-state logic

Three-state logic, also called TRI-STATE logic (a trademark of National Semiconductor Corporation, who invented it), provides an elegant solution. The name is misleading; it is not digital logic with three voltage levels. It's just ordinary logic,

with a third output state: open circuit (Fig. 8.19). A separate *enable* input determines whether the output behaves like an ordinary active pullup output or goes into the “third” (open) state, regardless of the logic levels present at the other inputs. Three-state outputs are available on many digital chips, including counters, latches, registers, etc., as well as on gates and inverters. A device with three-state output behaves exactly like ordinary active pullup logic when enabled, always driving its output either HIGH or LOW; when disabled, it effectively disconnects its output, so another logic device can drive the same line. Let’s look at an example.

A look ahead: data buses

Three-state drivers are widely used to drive computer data buses. Every device (memory, peripherals, etc.) that needs to put data on the (shared) bus ties onto it with three-state gates (or more complex functions such as “registers”). Things are cleverly arranged so that at most one device has its drivers enabled at any instant, all other devices being disabled into the open (third) state. In a typical situation the selected device “knows” to assert data onto the bus by recognizing its particular address on a set of address and control lines (Fig. 8.20). In this simplified

case the device is wired as port 6: It looks at address lines A_0 – A_2 and asserts data onto data bus $\bar{D}_0$ through $\bar{D}_3$ when it sees its particular address (i.e., 6) on the address lines *and* it sees a READ pulse. Such a bus protocol is adequate for many simple systems. Something like this is used in most microcomputers, as you will see in Chapters 10 and 11.

Note that there must be some external logic to make sure that three-state devices sharing the same output lines don’t try to talk at the same time (that undesirable condition is officially called “bus contention”). In this case all is well as long as each device responds to a unique address.

Open-collector logic

The predecessor to three-state logic was “open-collector” logic, which allows you to share a single line among the outputs of several drivers. An open-collector (or open-drain) output simply omits the active pullup transistor of the output stage (Fig. 8.21). The name “open-collector” is a good one. When you use such gates, you must supply an external pullup resistor somewhere. Its value isn’t critical; a small-value resistor gives increased speed and improved noise immunity, at the expense of increased power dissipation and loading

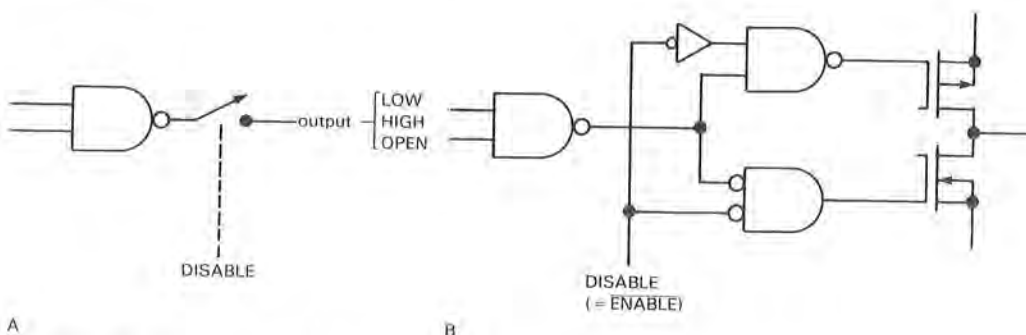


Figure 8.19. Three-state CMOS NAND gate.
A. Conceptual diagram.
B. Realization with internal CMOS gates.

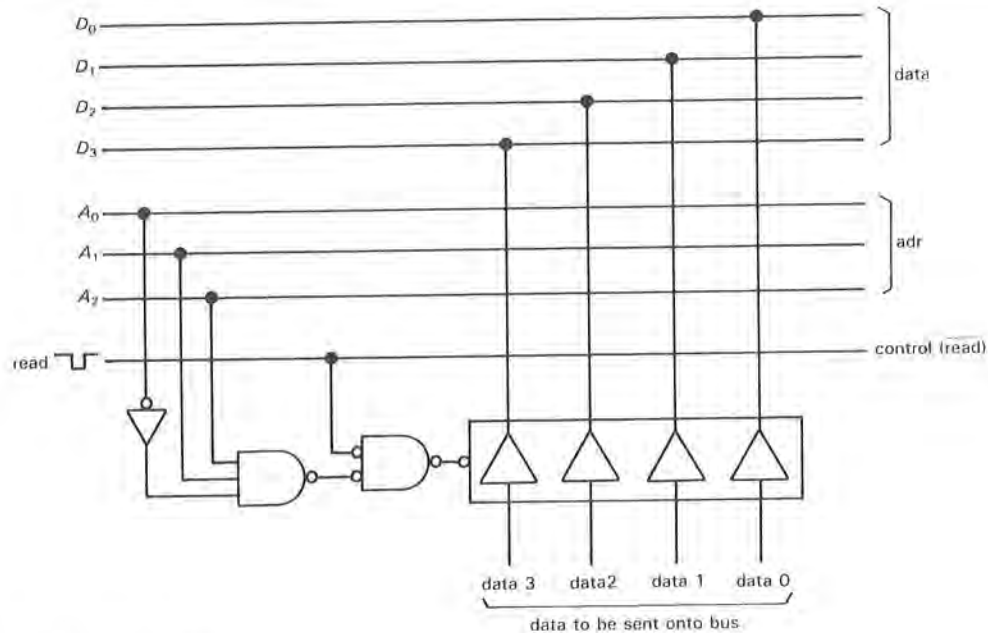


Figure 8.20. Data bus.

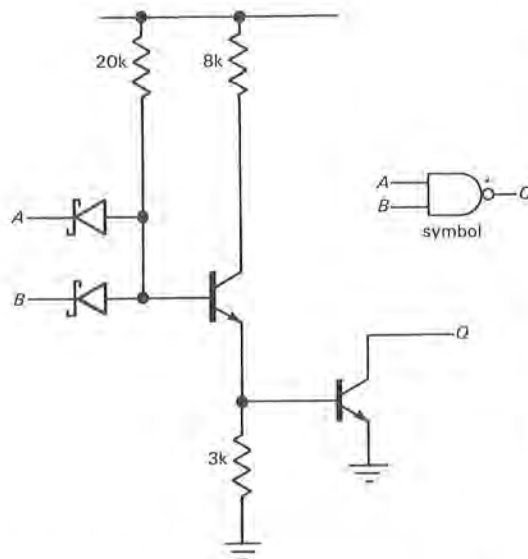


Figure 8.21. LS TTL open-collector NAND.

of the driver. Values of a few hundred to a few thousand ohms are typical. If you wanted to drive a bus with open-collector gates, you would substitute

2-input open-collector NANDs for the three-state drivers of Figure 8.20, bringing one input of each gate HIGH to enable the gates onto the bus; note that the data then asserted onto the bus are inverted. Each bus line would need a resistive pullup to +5 volts.

The disadvantage of open-collector logic is that speed and noise immunity are degraded, when compared with logic constructed with active pullup devices, because of the resistive pullup circuit. That's why three-state drivers are nearly universally favored for computer bus applications. However, there are three situations in which you would choose open-collector (or open-drain) devices: driving external loads, "wired-OR," and external buses. Let's look at them briefly.

Driving external loads

Open-collector logic is good for driving external loads that are returned to a higher-voltage positive supply. You might

want to drive a low-current lamp that requires 12 volts, or perhaps just generate a 15 volt logic swing by running a resistor from a gate's output to +15 volts, as in Figure 8.22. For example, the '06 is an open-collector hex inverter with 30 volt breakdown rating, and the CMOS 40107 is a dual-NAND open-drain buffer with up to 120mA sink capability. The 75450 series of "dual peripheral drivers" can sink up to 300mA from loads returned to +30 volts, and the UHP/UDN series from Sprague extends this to more than 1 amp and 80 volts. More on these subjects in the next chapter.

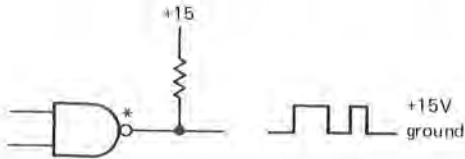


Figure 8.22

Wired-OR

If you wire together some open-collector gates as shown in Figure 8.23, you get what's called "wired-OR" – the combination behaves in this case like a larger NOR gate, with the output going LOW if any input is HIGH. You can't do this with active pullup outputs, because there would be a contest of wills, if all the gates didn't agree on what the output should be. You can combine NORs, NANDs, etc., with this kind of connection, and the output will be LOW if any gate asserts a LOW output. This connection is sometimes called "wired-AND," since the output is HIGH only if all gates have HIGH (open) outputs. Both names are describing the same thing: It's wired-AND for positive-true logic and wired-OR for negative-true logic. This will make more sense to you after you've seen DeMorgan's theorem, in the next section.

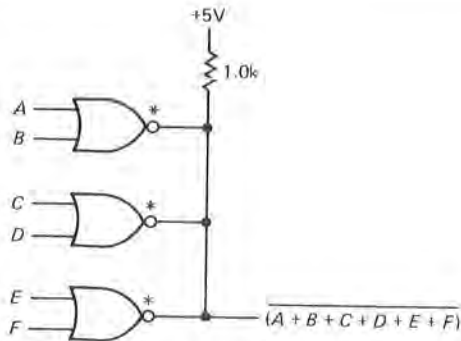


Figure 8.23. Wired-OR.

Wired-OR enjoyed some brief popularity in the early days of digital electronics, but it is not much used today, with two exceptions: (a) In the family of logic known as ECL (emitter-coupled logic) the outputs are what you might call "open-emitter," and can be wired-OR'ed painlessly, and (b) there are some shared lines in computer buses (most notably the line called *interrupt*) whose function is not to transfer data bits, but merely to indicate if *at least one* device is requesting attention; in that case you use wired-OR, since it does what you want and doesn't require external logic to prevent contention.

External buses

Where speed is not too important, you sometimes see open-collector drivers used to drive buses. This is often the case for buses that carry data out of computers; common examples are the buses used to connect to computer disk drives, and the IEEE-488 (also called "HPIB" or "GPIB") instrument bus. More on this in Chapters 10 and 11.

COMBINATIONAL LOGIC

As we discussed earlier in Section 8.04, digital logic can be divided into *combinational* and *sequential*. Combinational circuits are those in which the output state

depends only on the present input states in some predetermined fashion, whereas in sequential circuits the output state depends both on the input states and on the previous history. Combinational circuits can be constructed with gates alone, whereas sequential circuits require some form of memory (flip-flops). In these sections we will explore the possibilities of combinational logic before entering the turbulent world of sequential circuits.

8.12 Logic identities

No discussion of combinational logic is complete without the identities shown in Table 8.3. Most of these are obvious. The last two compose DeMorgan's theorem, the most important for circuit design.

TABLE 8.3. LOGIC IDENTITIES

$$\begin{aligned}
 ABC &= (AB)C = A(BC) \\
 AB &= BA \\
 AA &= A \\
 A1 &= A \\
 A0 &= 0 \\
 A(B+C) &= AB + AC \\
 A + AB &= A \\
 A + BC &= (A + B)(A + C) \\
 A + B + C &= (A + B) + C = A + (B + C) \\
 A + B &= B + A \\
 A + A &= A \\
 A + 1 &= 1 \\
 A + 0 &= A \\
 1' &= 0 \\
 0' &= 1 \\
 A + A' &= 1 \\
 AA' &= 0 \\
 (A')' &= A \\
 A + A'B &= A + B \\
 (A + B)' &= A'B' \\
 (AB)' &= A' + B'
 \end{aligned}$$

Example: exclusive-OR gate

We will illustrate the use of the identities with an example: making the exclusive-

OR function from ordinary gates. Figure 8.24 shows the XOR truth table. From studying this, and by realizing that the output is 1 only when $(A, B) = (0, 1)$ or $(1, 0)$, we can write

$$A \oplus B = \bar{A}B + A\bar{B}$$

from which we have the realization shown in Figure 8.25. However, this realization is not unique. Applying the identities, we find

$$\begin{aligned}
 A \oplus B &= A\bar{A} + A\bar{B} + B\bar{A} + B\bar{B} \\
 &\quad (A\bar{A} = B\bar{B} = 0) \\
 &= A(\bar{A} + \bar{B}) + B(\bar{A} + \bar{B}) \\
 &= A(\overline{AB}) + B(\overline{AB}) \\
 &= (A + B)(\overline{AB})
 \end{aligned}$$

A	B	$A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

Figure 8.24. XOR.

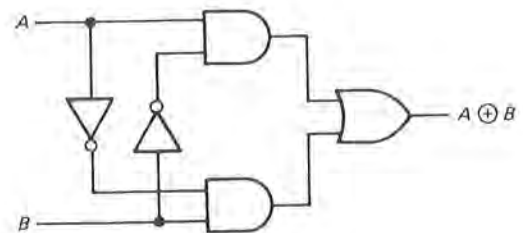


Figure 8.25. XOR realization.

(In the first step we used the trick of adding two quantities that equal zero; in the third step we used DeMorgan's theorem.) This has the realization shown in Figure 8.26. There are still other ways to construct XOR. Consider the following exercise.

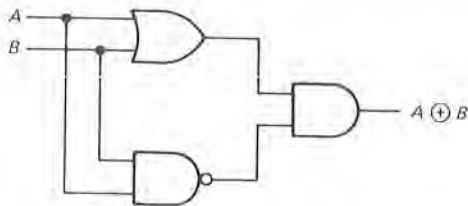


Figure 8.26. XOR realization.

EXERCISE 8.11

Show that

$$A \oplus B = \overline{AB} + \overline{A\overline{B}}$$

$$A \oplus B = (A + B)(\overline{A} + \overline{B})$$

by logic manipulation. You should be able to convince yourself that these are true by inspection of the truth table, combined with suitable hand-waving.

EXERCISE 8.12

What are the following: (a) $0 \cdot 1$, (b) $0 + 1$, (c) $1 \cdot 1$, (d) $1 + 1$, (e) $A(A + B)$, (f) $A(A' + B)$, (g) $A \text{ XOR } A$, (h) $A \text{ XOR } A'$?

8.13 Minimization and Karnaugh maps

Since a realization of a logic function (even one as simple as exclusive-OR) isn't unique, it is often desirable to find the simplest, or perhaps most conveniently constructed, circuit for a given function. Many good minds have worked on this problem, and there are several methods available, including algebraic techniques that can be coded to run on a computer. For problems with four or fewer inputs, a Karnaugh map provides one of the nicest methods; it also enables you to find a logic expression (if you don't know it) once you can write down the truth table.

We will illustrate the method with an example. Suppose you want to generate a logic circuit to count votes. Imagine that you have three positive-true inputs (each either 1 or 0) and an output (0 or 1). The

output is to be 1 if at least two of the inputs are 1.

Step 1. Make a truth table:

A	B	C	Q
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

All possible permutations must be represented, with corresponding output(s). Write an X (= "don't care") if either output state is OK.

Step 2. Make a Karnaugh map. This is somewhat akin to a truth table, but the variables are represented along two axes. Furthermore, they are arranged in such a way that only one input bit changes in going from one square to an adjacent square (Fig. 8.27).

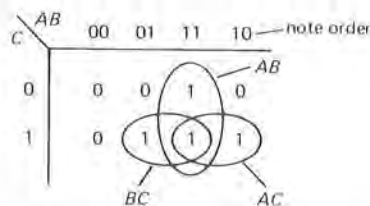


Figure 8.27. Karnaugh map.

Step 3. Identify on the map groups of 1's (alternatively, you could use groups of 0's): The three blobs enclose the logic expressions AB , AC , and BC . Finally, read off the required function, in this case

$$Q = AB + AC + BC$$

with the realization shown in Figure 8.28. The result seems obvious, in retrospect. We could have read off the pattern of 0's to get instead

$$Q' = A'B' + A'C' + B'C'$$

which might be useful if the complements A' , B' , and C' already exist somewhere in the circuit.

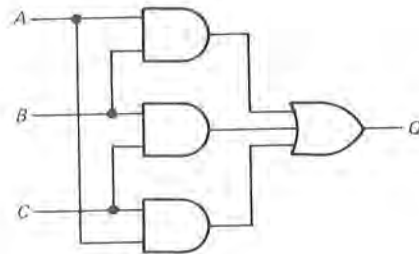


Figure 8.28

Some comments on Karnaugh maps

1. Look for groups of 2, 4, 8, etc., squares; they have the simple logic expressions.
2. The larger the block you describe, the simpler the logic.
3. The edges of the Karnaugh map connect up. For instance, the map in Figure 8.29 is described by $Q = B'C$.

AB \ CD	00	01	11	10
00	0	0	0	0
01	0	0	0	0
11	1	0	0	1
10	1	0	0	1

Figure 8.29

4. A block of 1's with only one or two 0's may be best described by the grouping illustrated in Figure 8.30, which corresponds to the logic expression

$$Q = A(BCD)'$$

5. Xs (don't care) are "wild cards." Use them as 1's or 0's to generate the simplest logic.

AB \ CD	00	01	11	10
00	0	0	1	1
01	0	0	1	1
11	0	0	0	1
10	0	0	1	1

Figure 8.30

6. A Karnaugh map may not lead directly to the best solution: A more complicated logic expression may sometimes have a simpler realization in gates, e.g., if some of its terms already exist as logic in your circuit, and you can exploit intermediate outputs (from other terms) as inputs. Furthermore, exclusive-OR realizations are not always obvious from Karnaugh maps. Finally, package constraints (e.g., the fact that four 2-input gates come in a single IC) also figure into the choice of logic used in the final circuit realization. When programmable logic devices such as PALs (Section 8.15) are used to construct logic functions, the internal structure (programmable AND; fixed OR) constrains the realization that can be used.

EXERCISE 8.13

Draw a Karnaugh map for logic to determine if a 3-bit integer (0 to 7) is prime (assume that 0, 1, and 2 are not primes). Show a realization with 2-input gates.

EXERCISE 8.14

Find logic to perform multiplication of two 2-bit unsigned numbers (i.e., each 0 to 3), producing a 4-bit result. Hint: Use a separate Karnaugh map for each output bit.

8.14 Combinational functions available as ICs

Using Karnaugh maps, you can construct logic to perform rather complicated functions such as binary addition or magnitude comparison, parity checking, multiplexing

(selecting one of several inputs, as determined by a binary address), etc. In the real world the most frequently used complex functions are available as single MSI functions (medium-scale integration, upward of 100 gates on one chip). Although many of the MSI functions involve flip-flops, which we will get to shortly, lots of them are combinational functions involving only gates. Let's see what animals live in the MSI combinational zoo.

Quad 2-input select

The quad 2-input select is a very useful chip. It is basically a 4-pole 2-position switch for logic signals. Figure 8.31 shows the basic idea. When SELECT is LOW, the A inputs are passed through to their respective Q outputs. For SELECT HIGH, the B inputs appear at the output. ENABLE' HIGH disables the device by forcing all outputs LOW. This is an important concept you will see more of later. Here's the truth table, which illustrates the X (don't care) entry:

E'	Inputs			Outputs Q_n
	SEL	A_n	B_n	
H	X	X	X	L
L	L	L	X	L
L	L	H	X	H
L	H	X	L	L
L	H	X	H	H

Figure 8.31 and the preceding table correspond to the '157 quad 2-input select chip. The same function is also available with inverted output ('158) and with 3-state outputs (true: '257; inverted: '258).

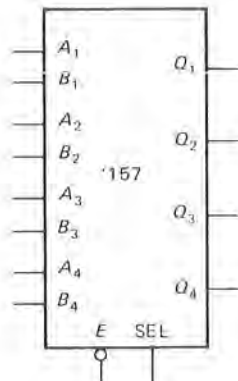


Figure 8.31. Quad 2-input select.

Although the function of a select gate can be performed by a mechanical switch in some cases, the gate is a far better solution, for several reasons: (a) it is cheaper; (b) all channels are switched simultaneously and rapidly; (c) it can be switched, nearly instantaneously, by a logic level generated elsewhere in the circuit; (d) even if the select function is to be controlled by a front-panel switch, it is better not to run logic signals around through cables and switches, to avoid capacitive signal degradation and noise pickup. With a select gate actuated by a dc level, you keep logic signals on the circuit board and get the bonus of simpler off-board wiring (a single line with pullup switched to ground by an SPST switch). Controlling circuit functions with externally generated dc levels in this manner is known as "cold switching," and it is a much better approach than controlling the signals themselves with switches, potentiometers, etc. Besides its other advantages, cold switching lets you bypass control lines with capacitors to eliminate interference, whereas signal lines cannot generally be bypassed. You will see some examples of cold switching later.

Transmission gates

As we discussed in Sections 3.11 and 3.12, with CMOS it is possible to make "transmission gates," simply a pair of

EXERCISE 8.15

Show how to make a 2-input select from an AND-OR-INVERT gate.

complementary MOSFET switches in parallel, so that an input (analog) signal between ground and V_{DD} is either connected through to the output through a low resistance (a few hundred ohms) or open-circuited (essentially infinite resistance). As you may remember, such a device is bidirectional and doesn't know (or care) which end is input and which end is output. Transmission gates work perfectly well with digital CMOS levels and are used extensively in CMOS design. Figure 8.32 shows the layout of the popular 4066 CMOS "quad bilateral switch." Each switch has a separate "control" input; input HIGH closes the switch, and LOW opens it. Note that transmission gates are merely switches, and therefore have no fanout; i.e., they simply pass input logic levels through to the output, without providing additional drive capability.

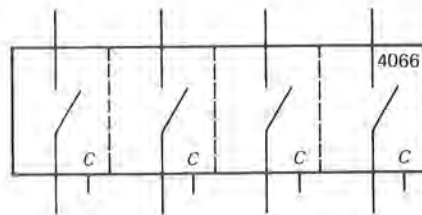


Figure 8.32. Quad transmission gate.

With transmission gates you can make 2-input (or more) select functions, usable with CMOS digital levels or analog signals. To select among a number of inputs, you can use a bunch of transmission gates (generating the control signals with a "decoder," as will be explained later). This is such a useful logic function that it has been institutionalized as the "multiplexer," which we will discuss next.

EXERCISE 8.16

Show how to make a 2-input select with transmission gates. You will need an inverter.

Multiplexers

The 2-input select gate is also known as a 2-input multiplexer. Multiplexers are also available with 4, 8, and 16 inputs (the 4-input variety comes as a dual unit, 2 in one package). A binary address is used to select which of the input signals appears at the output. For instance, an 8-input multiplexer has a 3-bit address input to address the selected data input (Fig. 8.33). The digital multiplexer illustrated is a '151. It has a STROBE (another name for ENABLE) input (negative-true), and it provides true and complemented outputs. When the chip is disabled (STROBE held HIGH), Q is LOW and Q' is HIGH, independent of the states of the address and data inputs.

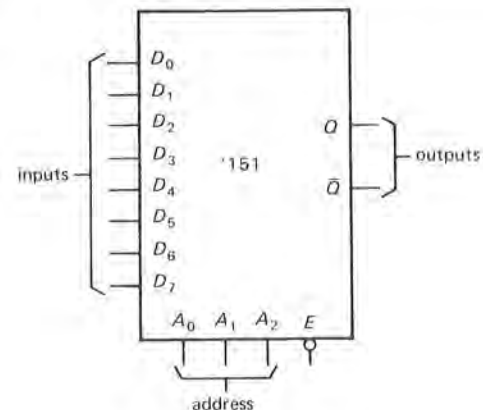


Figure 8.33. 8-input multiplexer.

In CMOS, two varieties of multiplexers are available. One type is for digital levels only, with an input threshold and "clean" regeneration of output levels according to the input state; that's also the way all TTL functions work. An example is the '153 TTL multiplexer. The other kind of CMOS multiplexer is analog and bidirectional; it's really just an array of transmission gates. The 4051–4053 CMOS multiplexers work this way (remember that logic made from transmission gates has no fanout). Since

transmission gates are bidirectional, these multiplexers can be used as “demultiplexers,” or decoders. We will discuss them next.

EXERCISE 8.17

Show how to make a 4-input multiplexer using (a) ordinary gates, (b) gates with 3-state outputs, and (c) transmission gates. Under what circumstances would (c) be preferable?

You might wonder what to do if you want to select among more inputs than are provided in a multiplexer. This question comes under the general category of chip “expansion” (using several chips that have small individual capabilities to generate a larger capability), and it applies to decoders, memories, shift registers, arithmetic logic, and many other functions as well. In this case the job is easy (Fig. 8.34). Here we have expanded two 74LS151 8-input multiplexers into a 16-input multiplexer. There’s an additional address bit, of course, and you use it to enable one chip or the other. The disabled chip holds its Q LOW, so an OR gate at the output completes the expansion. With three-state outputs the job is even simpler, since you can connect the outputs directly together.

Demultiplexers and decoders

A demultiplexer takes an input and routes

it to one of several possible outputs, according to an input binary address. The other outputs are either held in the inactive state or open-circuited, depending on the type of demultiplexer.

A decoder is similar, except that the address is the only input, and it is “decoded” to assert one of n possible outputs. Figure 8.35 shows an example. This is the ‘138 “1-of-8 decoder.” The output corresponding to (addressed by) the 3-bit input data is LOW; all others are HIGH. This particular decoder has three ENABLE inputs, all of which must be asserted (two LOW, one HIGH); otherwise all outputs are HIGH. A favorite use of the decoder is to cause different things to happen, depending on the state of a “counter” chip that drives it (more on this, soon). Decoders are commonly used when interfacing to microprocessor, to trigger different actions depending on the address; we’ll treat this subject in detail in Chapter 10. Another common use of a decoder is to enable a sequence of actions in turn, according to an advancing address given by the output of a binary *counter* (Section 8.25). A close cousin of the ‘138 is the ‘139, a dual 1-of-4 decoder with a single LOW-true enable per section. Figure 8.36 shows how to use a pair of ‘138 1-of-8 decoders to generate a 1-of-16 decoder. No external gates are necessary,

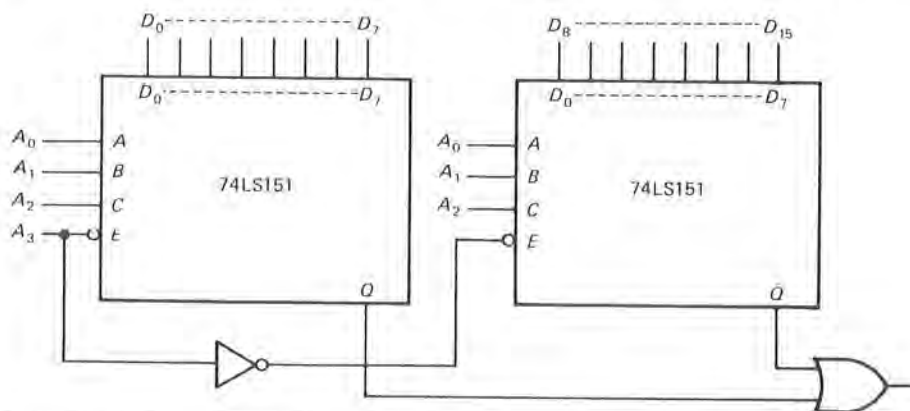


Figure 8.34. Multiplexer expansion.

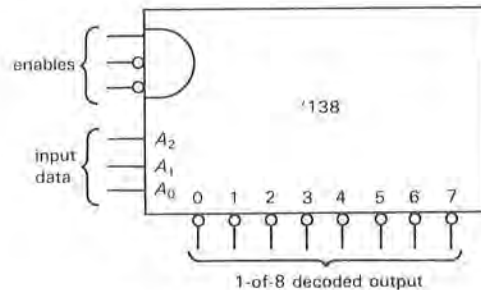


Figure 8.35. 1-of-8 decoder.

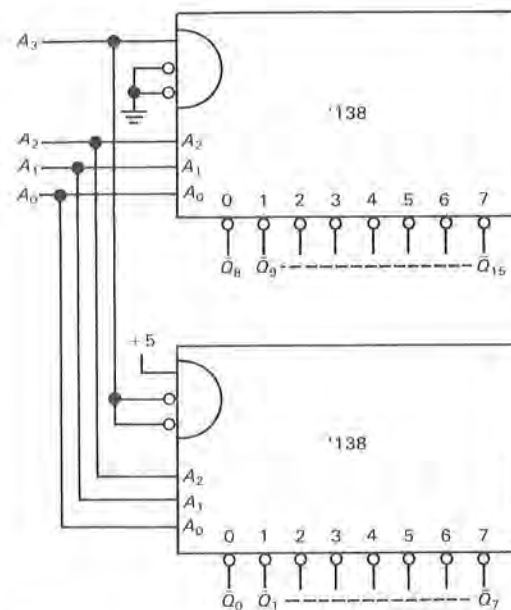


Figure 8.36. Decoder expansion.

since the '138 has enable inputs of both polarities.

EXERCISE 8.18

More expansion: Make a 1-of-64 decoder from nine '138s. Hint: Use one of them as an enabling switchyard for the others.

In CMOS logic, the multiplexers that use transmission gates are also demultiplexers, since transmission gates are bidirectional. When they are used that way, it is important to realize that the outputs that aren't selected are open-

circuited. A pullup resistor, or equivalent, must be used to assert a well-defined logic level on those outputs (the same requirement as with TTL open-collector gates).

There is another kind of decoder generally available in all logic families. An example is the '47 "BCD-to-7-segment decoder/driver." It takes a BCD input and generates outputs on 7 lines corresponding to the segments of a "7-segment display" that have to be lit to display the decimal character. This type of decoder is really an example of a "code converter," but in common usage it is called a decoder. Table 8.6 at the end of the chapter lists most available decoders.

EXERCISE 8.19

Design a BCD-to-decimal (1-of-10) decoder using gates.

Priority encoder

The priority encoder generates a binary code giving the address of the highest-numbered input that is asserted. It is particularly useful in "parallel-conversion" A/D converters (see next chapter) and in microprocessor system design. An example is the '148 8-input (3 output bits) priority encoder. The '147 encodes 10 inputs.

EXERCISE 8.20

Design a "simple" encoder: a circuit that outputs the (2-bit) address telling which of 4 inputs is HIGH (all other inputs must be LOW).

Adders and other arithmetic chips

Figure 8.37 shows a "4-bit full adder." It adds the 4-bit number A_i to the 4-bit number B_i , generating a 4-bit sum S_i plus carry bit C_o . Adders can be "expanded" to add larger numbers: The "carry-in" input C_i is provided to accept the carry out of the next lower adder. Figure 8.38 shows how you would add two 8-bit numbers.

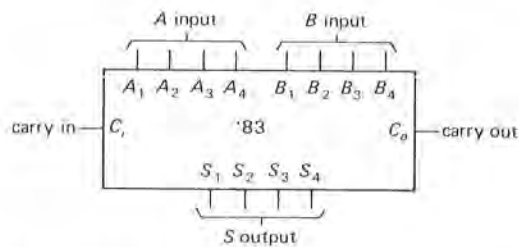


Figure 8.37. 4-bit full adder.

A device known as an arithmetic logic unit (ALU) is often used as an adder. It actually has the capability of performing a number of different functions. For instance, the '181 4-bit ALU (expandable to larger word lengths) can do addition, subtraction, bit shifts, magnitude comparison, and a few other functions. Adders and ALUs do their arithmetic in times measured in nanoseconds to tens of nanoseconds, depending on logic family.

Dedicated integer multiplier chips are available in configurations such as 8 bits times 8 bits, or 16 bits times 16 bits. A variation that is particularly suited to

digital signal processing is the so-called MAC (multiplier-accumulator), which accumulates a sum of products; these, too, are available in sizes up to 32×32 , with 64-bit product plus a few additional bits to keep the sum from overflowing. CMOS MACs and multipliers are available with typical speeds of 25–50ns; ECL multipliers are much faster – 5ns (typ) for 16×16 multiply.

Another arithmetic chip that's handy in digital signal processing is the *correlator*, which compares the corresponding bits of a pair of bit strings, calculating the number of bits that agree. A typical correlator chip compares a pair of 64-bit words, which can be shifted in internal shift registers. Any pattern of bits can be ignored ("masked") in the correlation. Typical speeds are 30ns, i.e., a bit stream can be clocked through at 35MHz, with a 7-bit correlation available at each clock tick. A variation (known as an FIR digital filter) calculates instead the true sum (with carry) of the true pairwise product of a pair of integer strings; typical

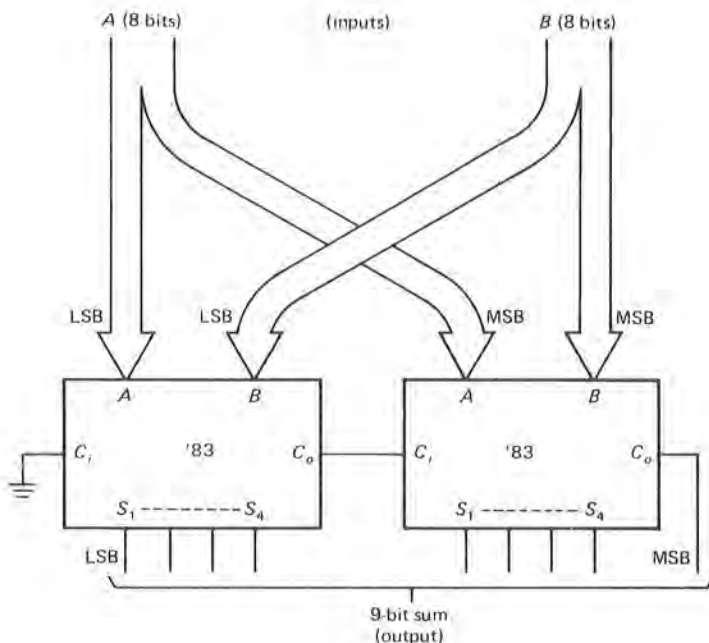


Figure 8.38. Adder expansion.

sizes are 4- to 10-bit integers, with string lengths of 3 to 8 words (expandable to greater length, of course).

The most complex arithmetic chips are the *floating-point processors*, which perform compares, sums, and products, as well as trigonometric functions, exponentials, and square roots. These are usually designed to work closely with particular microprocessors, and they generally conform to a nice standard known as IEEE P754, which specifies word size (up to 80 bits), format, etc. Examples (plus *matching microprocessors*) are the 8087 (8086/8), 80387 (80386), and 68881 (68020). These are truly stunning performers, with speeds of 10Mflops (million floating-point operations per second) or more.

Magnitude comparators

Figure 8.39 shows a 4-bit magnitude comparator. It determines the relative sizes of the 4-bit input numbers A and B and tells you via outputs whether $A < B$, $A = B$, or $A > B$. Inputs are provided for expansion to numbers larger than 4 bits.

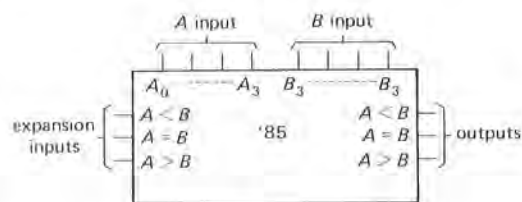


Figure 8.39. Magnitude comparator.

EXERCISE 8.21

Construct a magnitude comparator, using XOR gates, that tells whether or not $A = B$, where A and B are 4-bit numbers.

Table 8.7 at the end of the chapter lists most available magnitude comparators.

Parity generator/checker

This chip is used to generate a parity bit to be attached to a “word” when transmitting (or recording) data and to check the received parity when such data are recovered. Parity can be even or odd (e.g., with odd parity the number of 1 bits in each character is odd). The ’280 parity generator, for instance, accepts a 9-bit input word, giving an even and an odd parity bit output. The basic construction is an array of exclusive-OR gates.

EXERCISE 8.22

Figure out how to make a parity generator using XOR gates.

Programmable logic devices

You can build your own custom combinational (and even sequential) logic on a chip, using ICs that contain an array of gates with programmable interconnections. There are several varieties, of which the most popular are PALs (programmable array logic) and PLAs (programmable logic arrays). PALs, in particular, have become extremely inexpensive and flexible and should form a part of every designer’s toolbox of tricks. We will describe combinational PALs in the next section.

□ Some other strange functions

There are many other interesting MSI combinational chips worth knowing about. For example, in CMOS you can get a “majority logic” IC that tells you whether or not a majority of n inputs are asserted. Also available is a BCD “9’s complementer,” whose function is obvious. A “barrel-shifter” IC shifts an input word over by n (selectable) bits and can be expanded to any width.

8.15 Implementing arbitrary truth tables

Luckily, most of digital circuit design does not consist of cooking up crazy arrangements of gates to implement some complex logic function. However, there are times when you do need to wire up some complicated truth table, and the number of gates can become awfully large. You may begin to ask yourself if there isn't some other way. Fortunately, there are several. In this section we will look briefly at the use of multiplexers and demultiplexers to implement arbitrary truth tables. Then we will discuss the generally more powerful methods using programmable logic chips, particularly ROMs and PALs.

□ Multiplexers as generalized truth tables

It should be obvious that an n -input multiplexer can be used to generate any n -entry truth table, without any external components, by simply connecting its inputs to HIGH or LOW as required. For example, Figure 8.40 shows a circuit that tells if a 3-bit binary input is prime.

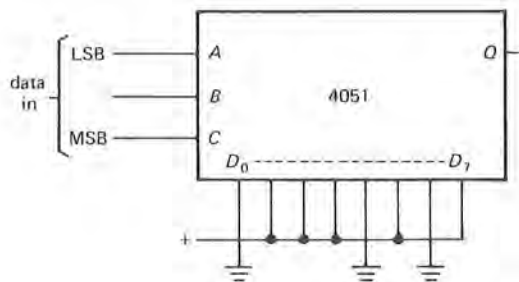


Figure 8.40

What is not so obvious is that an n -input multiplier can be used to generate any $2n$ -entry truth table, with at most one external inverter. For example, Figure 8.41 shows a circuit that tells whether or not a given month of the year has 31 days, where the month (1 to 12) is specified by a 4-bit input. The trick is to notice that for a given state of address bits applied to

the multiplexer, the output (as a function of the remaining input bit B) must be one of the four choices H, L, B , or B' ; the corresponding multiplexer input is therefore tied to logic HIGH, logic LOW, B or inverted B .

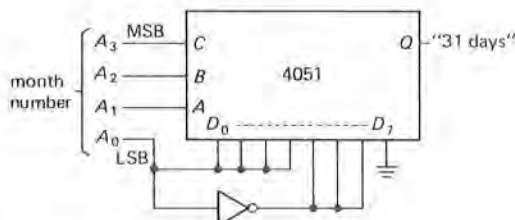


Figure 8.41

EXERCISE 8.23

Design of Figure 8.41. Make a table showing whether or not a given month has 31 days, with the month addressed in binary. Group the months in pairs, according to the most significant 3 bits of address. For each pair, figure out how Q ("31-ness") depends on the least significant address bit A_0 . Compare with Figure 8.41. Finally, verify (using your knuckles, or by reciting poetry) that the circuit does indeed tell you if a given month has 31 days.

Amusing postscript: It turns out that this truth table can be implemented with a single XOR gate, if you take advantage of X s (don't care) for the months that don't exist! Try your hand at this challenge. It will give you a chance to exercise Karnaugh map skills.

□ Decoders as generalized truth tables

Decoders also provide a nice shortcut for combinational logic, particularly in situations where you need several simultaneous outputs. As an example, let's generate a circuit to convert BCD to excess-3. Here's the truth table:

Decimal	BCD	XS3	Decimal	BCD	XS3
0	0000	0011	5	0101	1000
1	0001	0100	6	0110	1001
2	0010	0101	7	0111	1010
3	0011	0110	8	1000	1011
4	0100	0111	9	1001	1100

We use the 4-bit (BCD) input as an address to the decoder, then use the (negative-true) decoded outputs as inputs to several OR gates, one for each output bit, as shown in Figure 8.42. Note that with this scheme the output bits don't have to be mutually exclusive. You might use something like this as a cycle controller for a washing machine, in which you turn on several functions (pump out water, fill, spin, etc.) at each input state. You will see shortly

how to generate equally timed consecutive binary codes. The individual outputs from the decoder are known as *minterms*, and they correspond to positions on a Karnaugh map.

ROM and programmable logic

These are ICs that let you program their internal connections, roughly speaking. In that sense they are really devices with memory and should probably be discussed later, along with flip-flops, registers, etc. However, once programmed they are strictly combinational (although there are also sequential programmable logic devices;

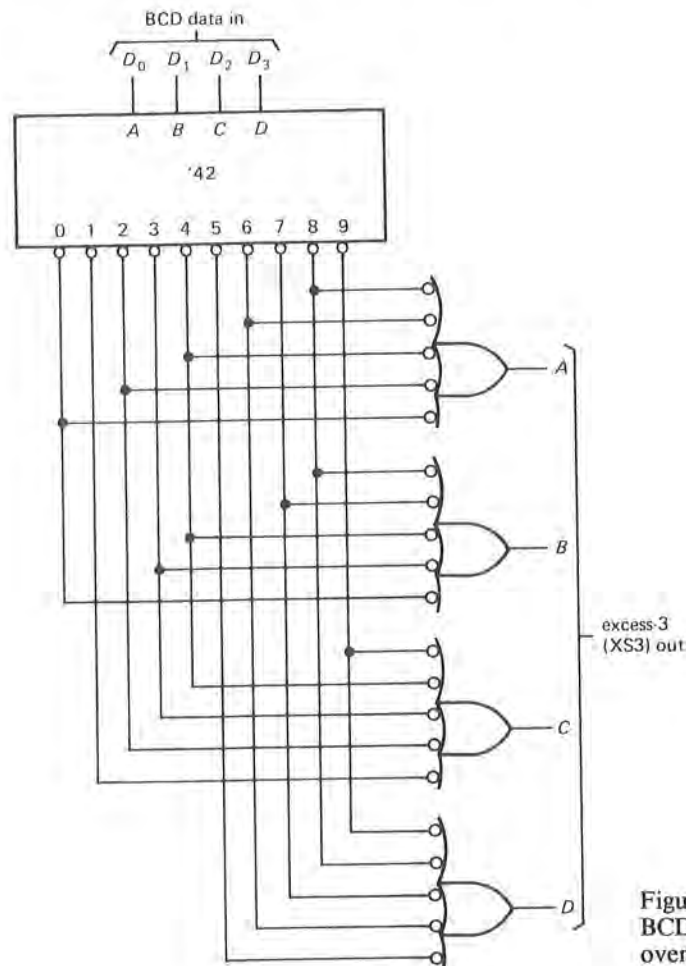


Figure 8.42. Minterm code conversion: BCD to "excess-3" (an obsolete code, left over from the first edition).

see Section 8.27), and they are so useful that it would be unforgivable not to discuss them now.

ROM. A ROM, or read-only memory, holds a bit pattern (typically 4 or 8 bits, parallel output) for each distinct address applied to its input. For example, a $1K \times 8$ ROM gives eight output bits for each of 1024 input states, specified by a 10-bit input address (Fig. 8.43). Any combinational truth table can therefore be programmed into a ROM, provided there are enough input (address) lines. For example, the $1K \times 8$ ROM above could be used to implement a 4-bit by 4-bit multiplier; in this case the limitation is the “width” (8 bits), not the “depth” (10 bits).

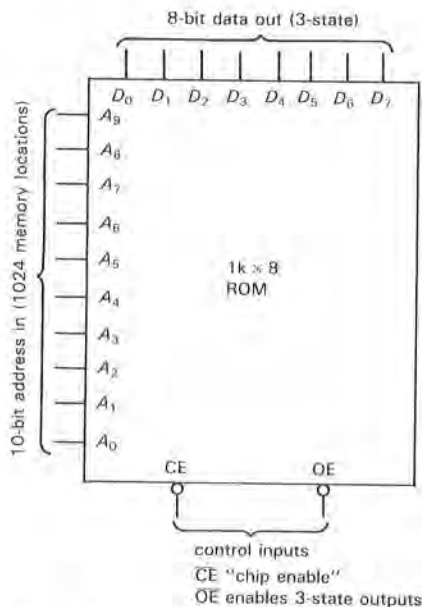


Figure 8.43

ROMs (and also programmable logic devices) are *nonvolatile*, meaning that the stored information is retained even when power is removed. There are several basic varieties, according to their method of programming: (a) “Mask-programmed ROMs” have their bit pattern built in at the time of manufacture. (b) “Program-

mable ROMs” (PROMs) are programmable by the user: PROMs use tiny interconnections that can be blown (like fuses) by applying appropriate address and control signals; they tend to be very fast (25–50ns), of relatively high power (bipolar: 0.5–1W), and small to medium in size (32×8 up to $8K \times 8$). (c) “Erasable programmable ROMs” (EPROMs) store their bits as charge held on floating MOS gates, and hence can be erased by exposing them to intense ultraviolet (UV) light for some tens of minutes (they have a transparent quartz window); they are available in NMOS and CMOS varieties and are rather slow (200ns), of low power (particularly in standby mode), and large ($8K \times 8$ to $128K \times 8$). Recent CMOS EPROMs are approaching bipolar speeds (35ns). A variant known as “one-time-programmable” (OTP) is an identical chip, but omits the quartz window for economy and ruggedness. (d) “Electrically erasable programmable ROMs” (EEPROMs) behave like EPROMs, but can be programmed and erased electrically, while in the circuit, with standard supply voltages (+5V).

ROMs find extensive use in computer and microprocessor applications, where they are used to store finished programs and data tables; we will see them again in Chapter 11. However, you should keep the smaller ROMs in mind as replacements for complicated arrays of gates.

Programmable logic. PALs (programmable array logic; PAL is a trademark of Monolithic Memories Inc.) and PLAs (programmable logic arrays) are the two basic kinds of programmable logic. They are ICs with many gates whose interconnections can be programmed (like ROMs) to form the desired logic functions. They are available in both bipolar and CMOS construction, the former using fusible-link (one-time-programmable), and the latter floating-gate MOS (UV or electrically erasable). You can’t program any

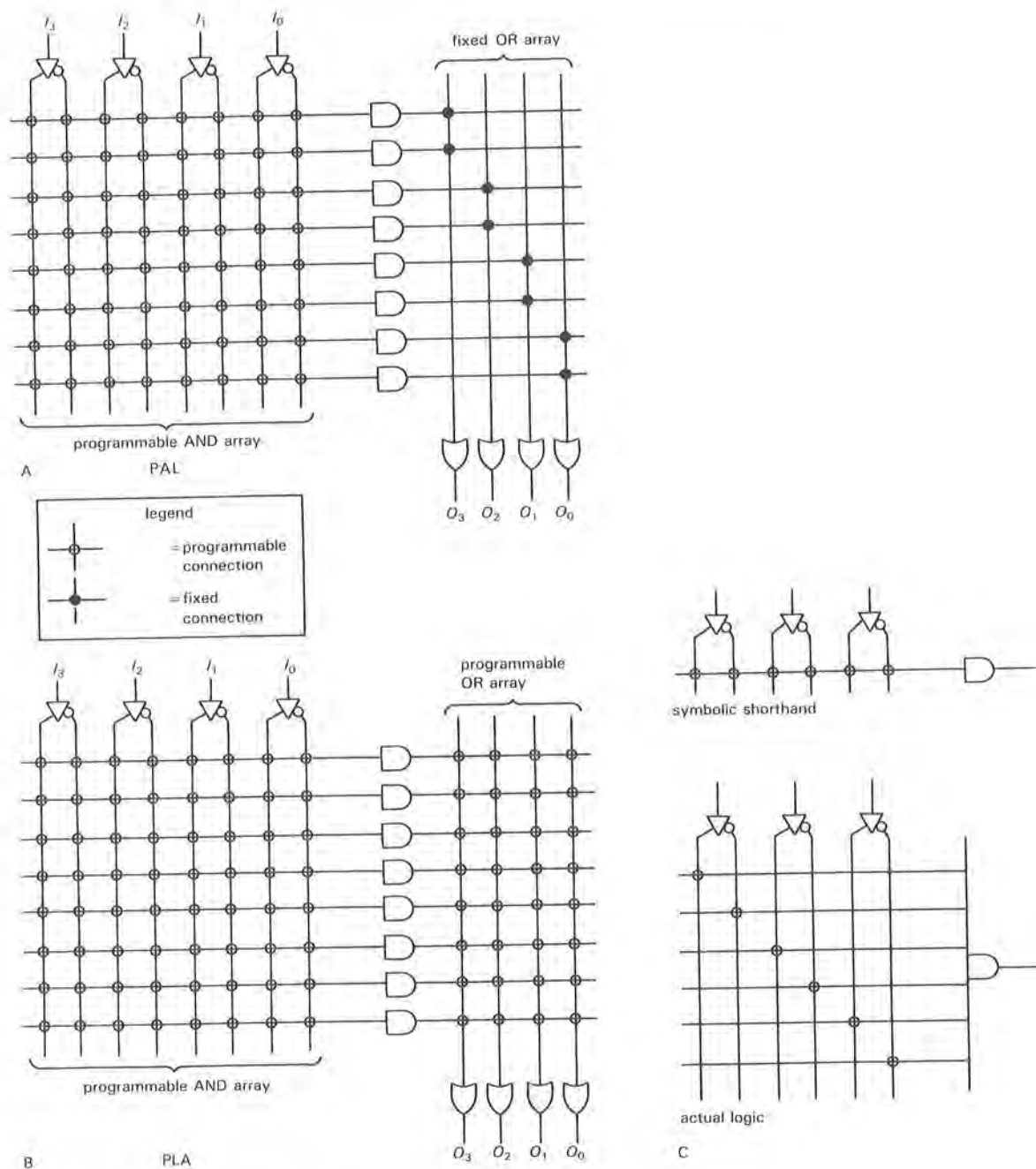


Figure 8.44. Programmable logic.

A. PAL

B. PLA

C. Detail of programmable connections to many-input AND gate; the circles are fusible links or other programmable connections.

interconnection you want – you’re limited by the built-in structure. Figure 8.44 shows the basic design of combinational (no registers) PALs and PLAs. To keep this figure simple, the AND and OR gates, though drawn with a single input line, are actually multiple-input gates, with an input implied at every connected crossing.

Each (three-state) output of a combinational PAL comes from an OR gate, each of whose inputs is prewired to an AND gate with dozens of inputs. For example, the 16L8 (Fig. 8.45) has eight 7-input OR gates; every possible signal is available at each AND gate, including the 10 dedicated input pins (and their inverts) and the 8 outputs (and their inverts). Each tristate enable is also derived from a 32-input AND gate.

PLAs are similar to PALs, but with the added flexibility that the AND-gate outputs can be connected to the OR-gate inputs in any combination (i.e., programmable), rather than being prewired as in a PAL.

Note that the PALs and PLAs that we have described are *combinational* (i.e., gates only, no memory). Both kinds of programmable logic are also available as *sequential* logic, i.e., with memory (registers), a subject we will take up in the next section.

To use PALs or PLAs, you get yourself a *programmer*, a piece of hardware that knows how to burn fuses (or otherwise program the device) and verify the finished product. All programmers connect via a serial port to a microcomputer (engineers have standardized on the IBM PC or compatible), on which you run some form of programmer software. Some of the fancier programmers include an on-board computer that runs its own software. The simplest kind of software simply lets you select the fuses to burn; you figure that out by deciding what logic you want, at the gate level, then listing (or marking on a graphics display) those fuses.

Figure 8.46 shows a trivial example, forming an exclusive-OR of two inputs as one of the outputs. Better programmers let you specify Boolean expressions (if you know them) or truth tables; the software does the rest, including minimization, simulation, and programming.

Although PLAs are more flexible, the overwhelming favorite in recent design has been the PAL. That is because they are faster (the signal passes through only one array of fuses) and cheaper and will usually do the job. As we’ll see shortly, sophisticated new PALs using “macrocells” and “folded architecture” give you some additional flexibility within the fixed-OR PAL design. PALs provide a flexible and compact alternative to fixed-function ICs and should not be overlooked by the serious circuit designer. We’ll show how (and when) to use programmable logic, along with useful tricks, in Section 8.27.

SEQUENTIAL LOGIC

8.16 Devices with memory: flip-flops

All our work with digital logic so far has been with combinational circuits (e.g., arrays of gates), for which the output is determined completely by the existing state of the inputs. There is no “memory,” no history, in these circuits. Digital life gets really interesting when we add devices with memory. This makes it possible to construct counters, arithmetic accumulators, and circuits that generally do one interesting thing after another. The basic unit is the flip-flop, a colorful name to describe a device that, in its simplest form, looks as shown in Figure 8.47.

Assume that both A and B are HIGH. What are X and Y ? If X is HIGH, then both inputs of G_2 are HIGH, making Y LOW. This is consistent with X being HIGH, so we’re finished. Right?

$X = \text{HIGH}$

$Y = \text{LOW}$

Logic Diagram

16L8

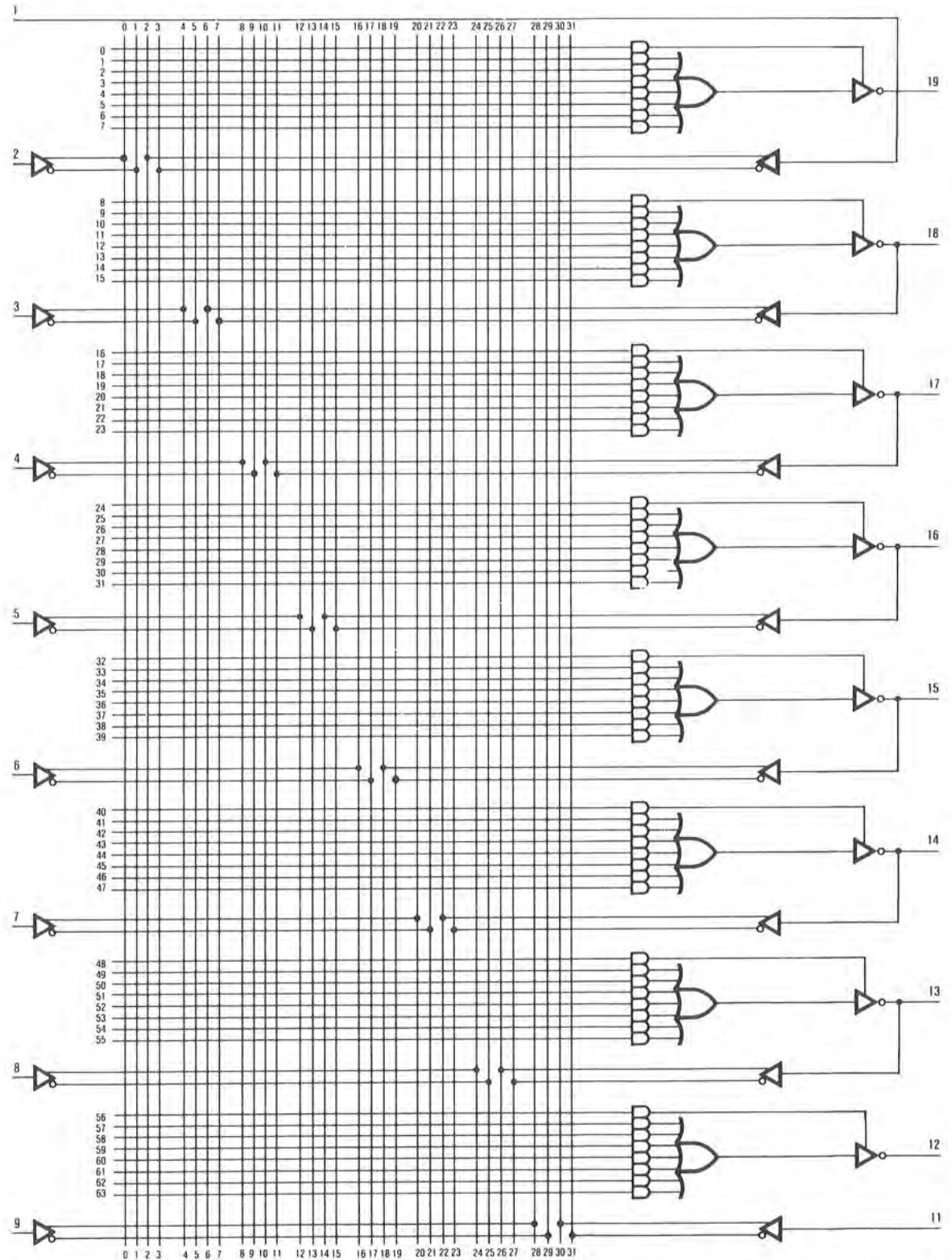


Figure 8.45. The 16L8 combinational PAL® has 10 dedicated inputs, 2 dedicated outputs, and 6 bidirectional (three-state) input/output lines; “16L8” means 16 (max) inputs, 8 (max) outputs (LOW-true). (Diagrams courtesy of Advanced Micro Devices of Sunnyvale, California.)

505

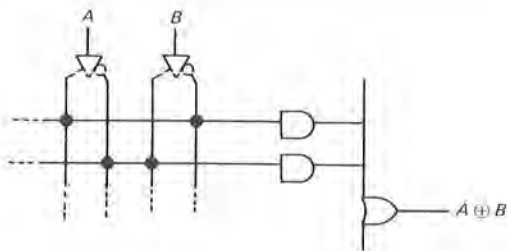


Figure 8.46. PAL exclusive OR.

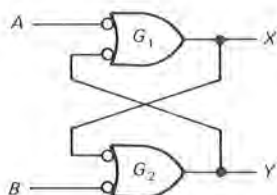


Figure 8.47. Flip-flop ("set-reset" type).

Wrong! The circuit is symmetrical, so an equally good state is

$X = \text{LOW}$

$Y = \text{HIGH}$

The states X, Y both LOW and X, Y both HIGH are not possible (remember, $A = B = \text{HIGH}$). So the flip-flop has two stable states (it's sometimes called a "bistable"). Which state it is in depends on past history. It has memory! To write into the memory, just bring one of the inputs momentarily LOW. For instance, bringing A LOW momentarily guarantees that the flip-flop goes into the state

$X = \text{HIGH}$

$Y = \text{LOW}$

no matter what state it was in previously.

Switch debouncing

This kind of flip-flop (with a SET and RESET input) is quite useful in many applications. Figure 8.48 shows a typical example. This circuit is supposed to enable the gate and pass input pulses when

the switch is opened. The switch is tied to ground (not +5V), because of a peculiarity of bipolar TTL (as opposed to CMOS): You must *sink* substantial current from an input in the LOW state (0.25mA for LS TTL), whereas in the HIGH state the input current is near zero. Besides, ground is generally available as a convenient return for switches and other controls. The problem with this circuit is that switch contacts "bounce." When the switch is closed, the two contacts actually separate and reconnect, typically 10 to 100 times over a period of about 1ms. You would get waveforms as sketched; if there were a counter or shift register using the output, it would faithfully respond to all those extra "pulses" caused by the bounce.

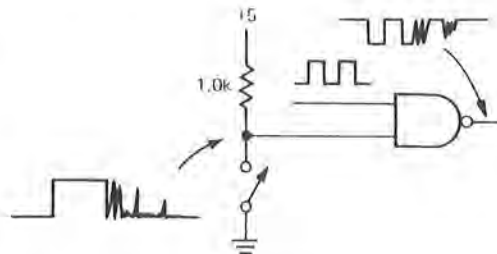


Figure 8.48. Switch "bounce."

Figure 8.49 shows the cure. The flip-flop changes state when the contacts first close. Further bouncing against that contact makes no difference (SPDT switches never bounce all the way back to the opposite position), and the output is a "debounced" signal, as sketched. This debouncer circuit is widely used; the '279 "quad SR latch" lets you get four into one package. Incidentally, the preceding circuit has a minor flaw: The first pulse after the gate is enabled may be shortened, depending on when the switch is closed relative to the input pulse train; the same holds for the final pulse of a sequence (of course, a switch that is not debounced has the same problem). A "synchronizer"

circuit (see Section 8.19) can be used to prevent this from happening, for applications where it makes a difference.

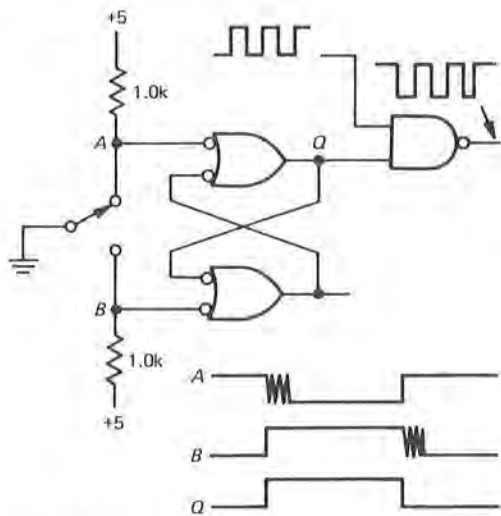


Figure 8.49. Switch debouncer.

Multiple-input flip-flop

Figure 8.50 shows another simple flip-flop. Here NOR gates have been used; a HIGH input forces the corresponding output LOW. Multiple inputs allow various signals to set or clear the flip-flop. In this circuit fragment, no pullups are used, since logic signals generated elsewhere (by standard active pullup outputs) are used as inputs.

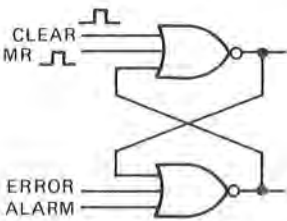


Figure 8.50

8.17 Clocked flip-flops

Flip-flops made with two gates, as in Figures 8.47 and 8.50, are known generically

as *SR* (set-reset), or jam-loaded, flip-flops. You can force them into one state or the other whenever you want by just generating the right input signal. They're handy for switch debouncing and many other applications. But the most widely used form of flip-flop looks a little different. Instead of a pair of jam inputs, it has one or two "data" inputs and a single "clock" input. The outputs can change state or stay the same, depending on the levels at the data inputs when the clock pulse arrives.

The simplest clocked flip-flop looks as shown in Figure 8.51. It's just our original flip-flop, with a pair of gates (controlled by the clock) to enable the SET and RESET inputs. It is easy to verify that the truth table is

<i>S</i>	<i>R</i>	<i>Q_{n+1}</i>
0	0	<i>Q_n</i>
0	1	0
1	0	1
1	1	indeterminate

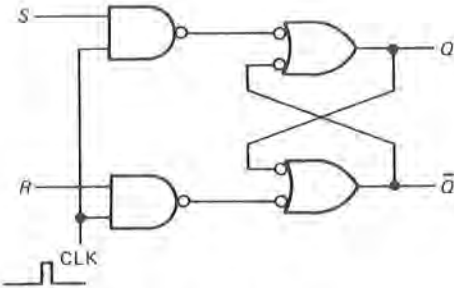


Figure 8.51. Clocked flip-flop.

where Q_{n+1} is the Q output after the clock pulse and Q_n is the output before the clock pulse. The basic difference between this and the previous flip-flops is that R and S should now be thought of as data inputs. What is present on R and S when a clock pulse comes along determines what happens to Q .

This flip-flop has one awkward property, however. The output can change in response to the inputs during the time the

clock is HIGH. In that sense it is still like the jam-loaded *SR* flip-flop (it's also known as a "transparent latch," since the output "sees through" to the input when the clock is HIGH). The full utility of clocked flip-flops comes with the introduction of slightly different configurations, the master-slave flip-flop and the edge-triggered flip-flop.

Master-slave and edge-triggered flip-flops

These are by far the most popular flip-flops. The data present on the input

lines just before a clock transition, or "edge," determines the output state after the clock has changed. These flip-flops are available as inexpensive packaged ICs and are always used in that form. But it is worth looking at their innards in order to understand what is going on. Figure 8.52 shows the schematics. Both are known as type *D* flip-flops. Data present at the *D* input will be transferred to the *Q* output after the clock pulse. The master-slave configuration is probably easier to understand. Here's how it works:

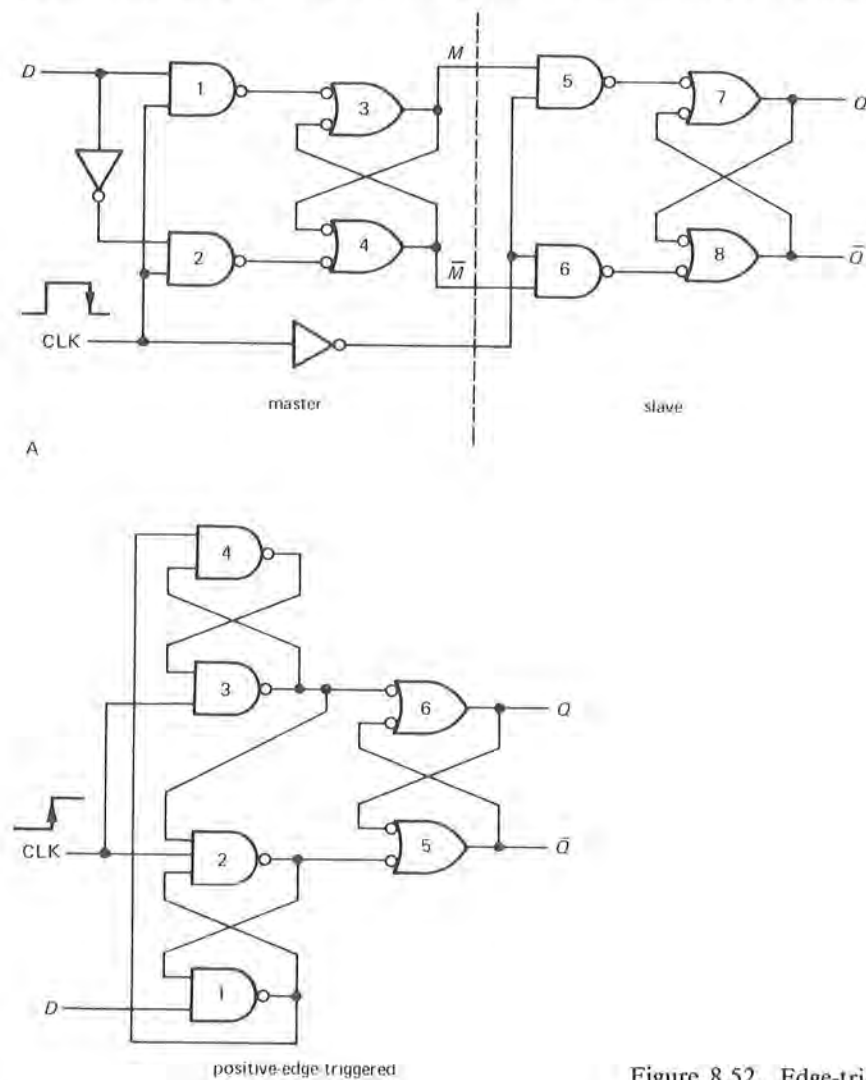


Figure 8.52. Edge-triggered type *D* flip-flops.

While the clock is HIGH, gates 1 and 2 are enabled, forcing the master flip-flop (gates 3 and 4) to the same state as the D input: $M = D, M' = D'$. Gates 5 and 6 are disabled, so the slave flip-flop (gates 7 and 8) retains its previous state. When the clock goes LOW, the inputs to the master are disconnected from the D input, while the inputs of the slave are simultaneously coupled to the outputs of the master. The master thus transfers its state to the slave. No further changes can occur at the output, because the master is now stuck. At the next rising edge of the clock, the slave will be decoupled from the master and will retain its state, while the master will once again follow the input.

The edge-triggered circuit behaves the same externally, but the inner workings are different. It is not difficult to figure it out. The particular circuit shown happens to be the popular '74 positive-edge-triggered type D flip-flop. The preceding master-slave circuit transfers data to the output on the negative edge. Flip-flops are available with either positive or negative edge triggering. In addition, most flip-flops also have SET and CLEAR jam-type inputs. They may be set and cleared on HIGH or on LOW, depending on the type of flip-flop. Figure 8.53 shows a few popular flip-flops. The wedge means "edge-triggered," and the little circles mean "negation," or complement. Thus, the '74 is a dual type D positive-edge-triggered flip-flop with active LOW jam-type SET and CLEAR inputs. The 4013 is a CMOS dual type D

positive-edge-triggered flip-flop with active HIGH jam-type SET and CLEAR inputs. The '112 is a dual JK master-slave flip-flop with data transfer on the negative edge and with active LOW jam-type SET and CLEAR inputs.

The JK flip-flop. The JK flip-flop works on principles similar to those of the type D flip-flop, but it has two data inputs. Here's the truth table:

J	K	Q_{n+1}
0	0	Q_n
0	1	0
1	0	1
1	1	Q_n'

Thus, if J and K are complements, Q will go to the value of the J input at the next clock edge. If J and K are both LOW, the output won't change. If J and K are both HIGH, the output will "toggle" (reverse its state after each clock pulse).

Warning: Some older JK flip-flops are "ones-catching," a term you won't find in the data sheet, but an effect that can have dire consequences for the unsuspecting. This means that if either J or K (or both) changes state momentarily while the slave is enabled by the clock, then returns to its previous state before the clock makes its transition, the flip-flop will "remember" that momentary state and behave as if that state had persisted. Thus, the flip-flop may change state at the next clock transition even if the J and K inputs existing at that transition should cause the flip-flop to remain in its current state. This can

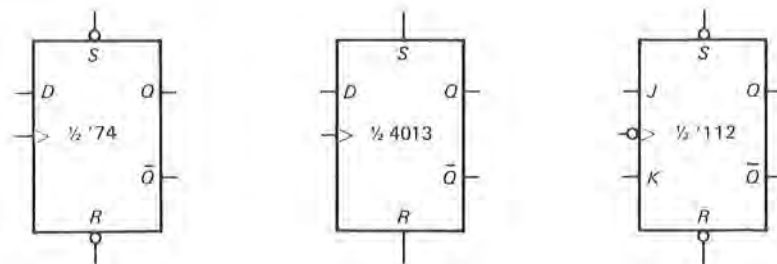


Figure 8.53. D -type and JK flip-flops.

lead to peculiar behavior, to put it mildly. The problem arises because such flip-flops were designed with short clock pulses in mind, whereas in common usage you clock flip-flops with just about anything. Be careful when using master-slave flip-flops, or avoid them altogether and use true edge-triggered flip-flops instead.

Two good choices that employ true edge triggering are the '112 and the '109. Both are dual (two per package) *JK* flip-flops with (negative-true) SET and CLEAR jam-type inputs; the '112 clocks on the negative edge, the '109 on the positive edge. The '109 has an interesting quirk, namely the *K* input is complemented (it's sometimes called a "*JK*-bar" flip-flop). Thus, if you tie the *J* and *K* inputs together, you've got a *D* flip-flop; to make it toggle, you ground *K'* and tie *J* HIGH.

Divide by 2

It is easy to make a divide-by-2 circuit by just exploiting the toggling capability of flip-flops. Figure 8.54 shows two ways. The *JK* flip-flop toggles when both inputs are HIGH, producing the output shown. The second circuit also toggles, since with the *D* input tied to its own *Q'* output, the *D* flip-flop always sees the complement of its existing output at its *D* input at the time of the clock pulse. The output signal in either case is at half the frequency of the input.

Data and clock timing

This last circuit raises an interesting question: Will the circuit fail to toggle, since the *D* input changes almost immediately after the clock pulse? In other words, will the circuit get confused, with such crazy things happening at its input? You could, instead, ask this question: Exactly *when* does the *D* flip-flop (or any other flip-flop) look at its input, relative to the clock pulse? The answer is that there is a specified "setup time" t_s and "hold time" t_h for any clocked device. Input data must

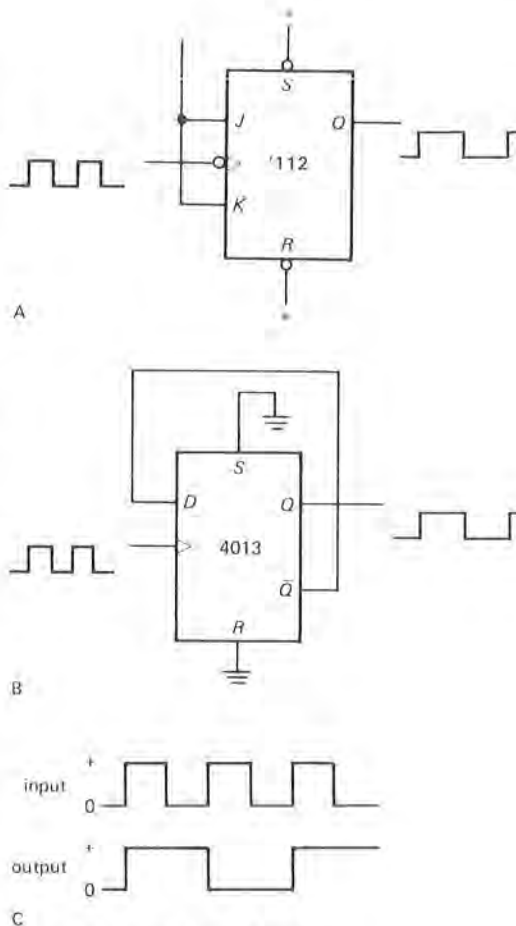


Figure 8.54. Toggling flip-flops.

be present and stable from at least t_s before the clock transition until at least t_h after it, for proper operation. For the 74HC74, for instance, $t_s = 20\text{ns}$ and $t_h = 3\text{ns}$ (Fig. 8.55). So, for the preceding toggling connection, the setup-time requirement is met if the output has been stable for at least 20ns before the next clock rising edge. It may look as if the hold-time requirement is violated, but that's OK, also. The minimum "propagation time" from clock to output is 10ns, so a *D* flip-flop connected to toggle as described is guaranteed to have its *D* input stable for at least 10ns after the clock transition. Most devices nowadays have a zero hold-time requirement.

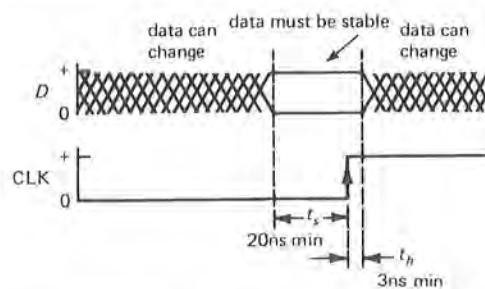


Figure 8.55. Data setup and hold times.

An interesting thing can happen if the level at the D input changes during the setup-time interval, namely a so-called *metastable* state in which the flip-flop can't make up its mind which state to go into. We will have more to say about this shortly.

Divide by more

By cascading several toggling flip-flops (connect each Q output to the next clock input), it is easy to make a divide-by- 2^n , or binary, counter. Figure 8.56 shows a four-stage "ripple counter" and its waveforms. Note that flip-flops that clock on the falling edge (indicated by the negation circle) must be used if each Q output drives the next clock input. This circuit is a divide-by-16 counter: The output waveform from the last flip-flop is a square wave whose frequency is $1/16$ of the circuit's input clock frequency. Such a circuit is called a *counter* because the data present at the four Q outputs, considered as a single 4-bit binary number, go through a binary sequence from 0 to 15, incrementing after each input pulse. The waveforms in Figure 8.56 demonstrate this fact. In the

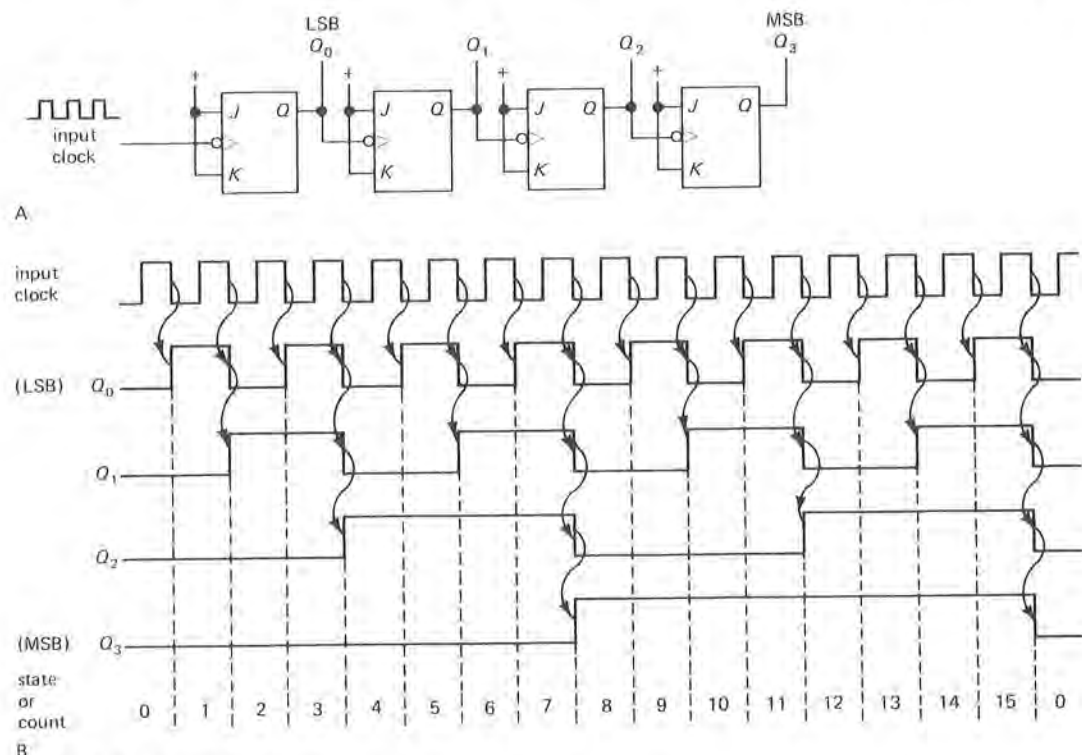


Figure 8.56. Four-bit counter.
A. Schematic.
B. Timing diagram.

figure the abbreviation MSB is used to mean “most significant bit,” and LSB means “least significant bit”; the curved arrows are used to indicate what causes what, to aid in understanding.

As you will see in Section 8.25, the counter is such a useful function that many versions are available integrated onto single chips, including 4-bit, BCD, and multidigit counting formats. By cascading several such counters and displaying the count on a numeric display device (e.g., an LED digital display) you can easily construct an event counter. If the input pulse train to such a counter is gated for exactly 1 second, you’ve got a frequency counter, which displays frequency (cycles per second) by actually counting the number of cycles in a second. Section 15.10 shows diagrams of this simple and highly useful scheme. In fact, single-chip frequency counters are available, complete with oscillator, counter, control, and display circuitry; see Figure 8.71 for an example.

In practice, the simple scheme of cascading counters by connecting each Q output to the next clock input has some interesting problems related to the cascaded delays as the signal “ripples” down through the chain of flip-flops, and a “synchronous” scheme (in which all clock inputs see the same clocking signal) is usually better. Let’s look into this question of synchronous clocked systems.

8.18 Combining memory and gates: sequential logic

Having explored the properties of flip-flops, let’s see what can be done when they are combined with the combinational (gate) logic we discussed earlier. Circuits made with gates and flip-flops constitute the most general form of digital logic.

Synchronously clocked systems

As we hinted in the last section, sequential logic circuits in which there is a common

source of clock pulses driving all flip-flops have some very desirable properties. In such a *synchronous system* all action takes place just after each clocking pulse, based on the levels present just before each clock pulse. Figure 8.57 shows the general scheme.

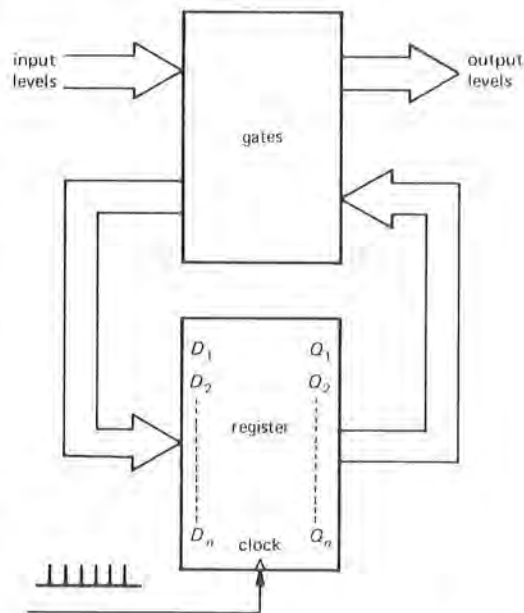


Figure 8.57. The classical sequential circuit: memory registers plus combinational logic. This scheme can be easily implemented with single-chip “registered PALs” (Section 8.27).

The flip-flops have all been combined into a single *register*, which is nothing more than a set of type D flip-flops with their clock inputs all tied together and their individual D inputs and Q outputs brought out; i.e., each clock pulse causes the levels present at the D inputs to be transferred to the respective Q outputs. The box full of gates looks at both the Q outputs and whatever input levels are applied to the circuit and generates a new set of D inputs and logic outputs. This simple-looking scheme is extremely powerful. Let’s look at an example.

Example: divide-by-3

Let's design a synchronous divide-by-3 circuit with two type *D* flip-flops, both clocked from the input signal. In this case, D_1 and D_2 are the register inputs, Q_1 and Q_2 are the outputs, and the common clock line is the master clocking input (Fig. 8.58).

1. Choose the three states. Let's use

Q_1	Q_2
0	0
0	1
1	0
0	0

(i.e., first state)

2. Find the combinational logic network outputs necessary to generate this sequence of states, i.e., figure out what the *D* inputs have to be to get those outputs:

Q_1	Q_2	D_1	D_2
0	0	0	1
0	1	1	0
1	0	0	0

3. Concoct suitable gating (combinational logic), using available outputs, to produce those *D* inputs. In general, you can use a Karnaugh map. In this simple case you can see by inspection that

$$D_1 = Q_2$$

$$D_2 = (Q_1 + Q_2)'$$

from which the circuit of Figure 8.59 follows.

It is easy to verify that the circuit works as planned. Since it is a synchronous counter, all outputs change simultaneously (when you feed one output to the next clock, you've got a *ripple* counter instead). In general, synchronous (or "clocked") systems are desirable, since susceptibility to noise is improved: Things have settled down by the time of the clock pulse, so circuits that only look at their inputs at clock edges aren't troubled by capacitively coupled interference from other flip-flops, etc. A further advantage of clocked systems is that transient states (caused by

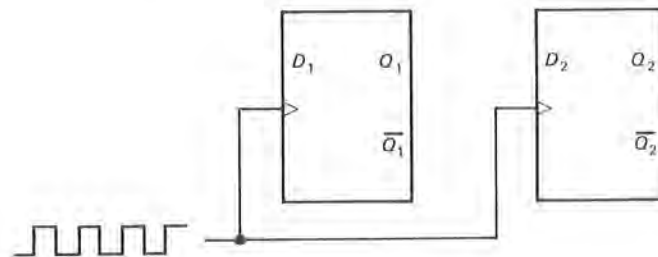


Figure 8.58

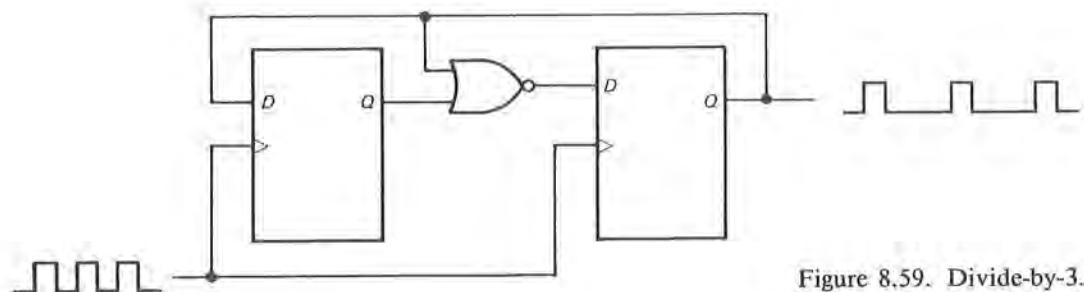


Figure 8.59. Divide-by-3.

delays, so that all outputs don't change simultaneously) don't produce false output, since the system is insensitive to what happens just *after* a clock pulse. You will see some examples later.

Excluded states

What happens to the divide-by-3 circuit if the flip-flops somehow get into the state $(Q_1, Q_2) = (1, 1)$? This can easily happen when the circuit is first turned on, since the initial state of a flip-flop is anyone's guess. From the diagram, it is clear that the first clock pulse will cause it to go to the state $(1, 0)$, from which it will function as before. It is important to check the excluded states of a circuit like this, since it is possible to be unlucky and have it get stuck in one of those states. (Alternatively, the initial design procedure can include a specification of all possible states.) A useful diagnostic tool is the *state diagram*, which for this example looks like Figure 8.60. Usually you write the conditions for each transition next to the arrows, if other variables of the system are involved. Arrows may go in both directions between states, or from one state to several others.

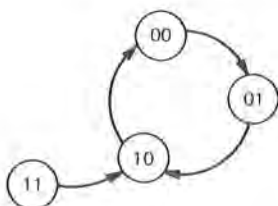


Figure 8.60. State diagram: divide-by-3.

EXERCISE 8.24

Design a synchronous divide-by-3 circuit using two *JK* flip-flops. It can be done (in 16 different ways!) without any gates or inverters. One

hint: When you construct the table of required J_1, K_1 and J_2, K_2 inputs, keep in mind that there are two possibilities for $\bar{J}, \bar{K}$ at each point. For instance, if a flip-flop output is to go from 0 to 1, $J, K = 1, X$ (X = don't care). Finally, check to see if the circuit will get stuck in the excluded state (of the 16 distinct solutions to this problem, 4 will get stuck and 12 won't).

EXERCISE 8.25

Design a synchronous 2-bit UP/DOWN counter: It has a clock input, and a control input (U/D'); the outputs are the two flip-flop outputs Q_1 and Q_2 . If U/D' is HIGH, it goes through a normal binary counting sequence; if LOW, it counts backward – $Q_2Q_1 = 00, 11, 10, 01, 00, \dots$

State diagrams as design tools

The state diagram can be very useful when designing sequential logic, particularly if the states are connected together by several paths. In this design approach, you begin by selecting a set of unique states of the system, giving each a name (i.e., a binary address). You will need a minimum of n flip-flops, or bits, where n is the smallest integer for which 2^n is equal to or greater than the number of distinct states in the system. Then you set down all the rules for moving between states, i.e., all possible conditions for entering and leaving each state. From there it is a straightforward (but perhaps tedious) job to generate the necessary combinational logic, since you have all possible sets of Q s and the set of D s that each leads to. Thus you have converted a sequential design problem into a combinational design problem, always soluble through techniques such as the Karnaugh map. Figure 8.61 shows a real-world example. Note that there may be states that don't lead to others, e.g., "receive diploma."

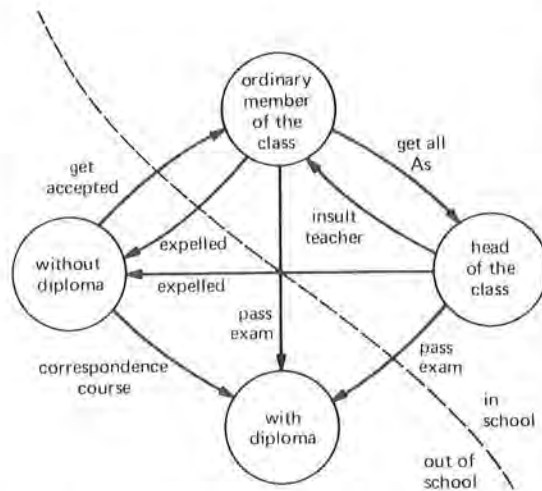


Figure 8.61. State diagram: going to school.

Registered PALs

Programmable logic (PALs and PLAs, see Section 8.15) is available with both gates and synchronously clocked *D* flip-flops on the same chip; these are known as *registered* PALs and PLAs, and they are ideal for implementing custom sequential circuits. We'll show how in Section 8.27.

8.19 Synchronizer

An interesting application of flip-flops in sequential circuits is their use in a *synchronizer*. Suppose you have some external control signal coming into a synchronous system that has clocks, flip-flops, etc., and you want to use the state of that input signal to control some action. For example, a signal from an instrument or experiment might signify that data are ready to be sent to a computer. Since the experiment and the computer march to the beats of different drummers, so to speak (in fancy language you would say they are *asynchronous* processes), you need a method to restore order between the two systems.

Example: pulse synchronizer

As an example, let's reconsider the circuit in which a debouncer flip-flop gated a pulse train (Section 8.16). That circuit enables the gate whenever the switch is closed, regardless of the phase of the pulse train being gated, so that the first or last pulse may be shortened. The problem is that the switch closure is asynchronous with the pulse train. In some applications it is important to have only *complete* clock cycles, and that requires a synchronizer circuit like that in Figure 8.62. Pushing START brings the output of gate 1 HIGH, but *Q* stays LOW until the next falling edge of the input pulse train. In that way, only complete pulses are passed by NAND gate 3. Figure 8.62 shows some waveforms. The curved arrows are drawn to show exactly what causes what. You can see, for instance, that the transitions of *Q* occur slightly *after* the falling edges of the input.

Logic races and glitches

This example brings up a subtle but extremely important point: What would happen if a positive-edge-triggered flip-flop were used instead? If you analyze it carefully, you'll find that START still works OK, but if STOP is pushed while the input is LOW, a bad thing happens (Fig. 8.63). A short spike, or "glitch," gets through because the final NAND gate isn't disabled until the flip-flop output has a chance to go LOW, a delay of about 20ns for HC or LS TTL. This is a classic example of a "logic race." With some care these situations can be avoided, as the example shows. Glitches are terrible things to have running through your circuits. Among other things, they're hard to see on an oscilloscope, and you may not know they are present. They can clock subsequent flip-flops erratically, and they may be widened – or narrowed to extinction – by passage through gates and inverters.

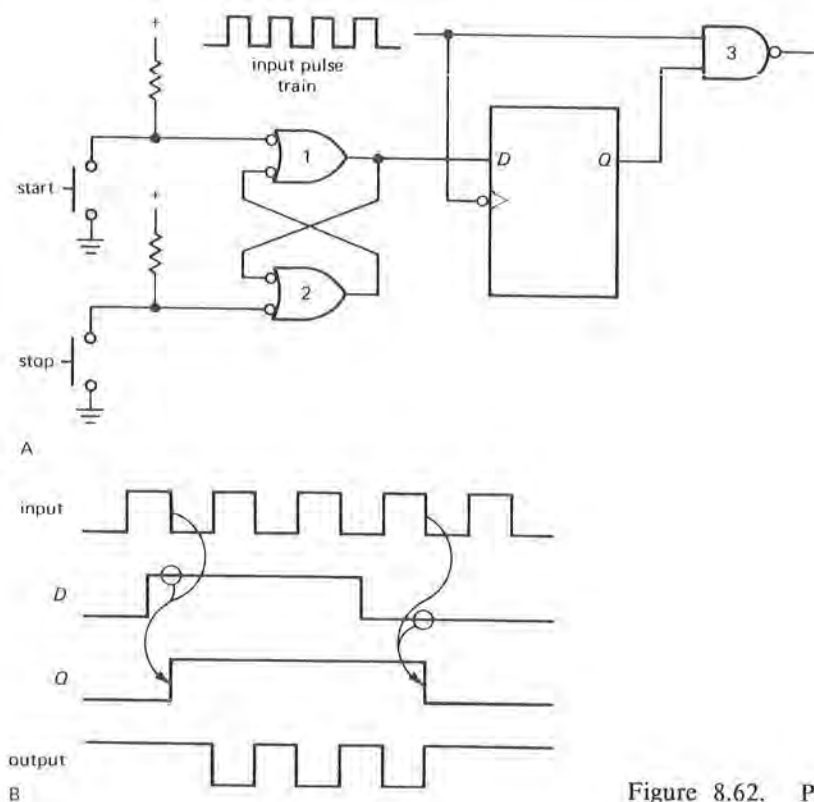


Figure 8.62. Pulse-train synchronizer.

EXERCISE 8.26

Demonstrate that the preceding pulse synchronizer circuit (Fig. 8.62) does not generate glitches.

EXERCISE 8.27

Design a circuit that lets exactly one full negative pulse (from an input train of pulses) pass through to the output, after a button is pushed.

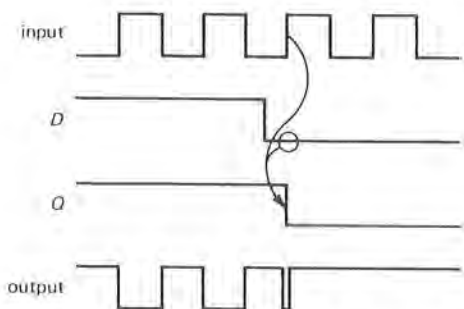


Figure 8.63. A logic race can generate a "runt pulse."

A few comments about synchronizers: The input to the *D* flip-flop can come from other logic circuitry, rather than from a debounced switch. There are applications in computer interfacing, etc., where an asynchronous signal must communicate with a clocked device; in such cases clocked flip-flops or synchronizers are ideal. In this circuit, as in all logic, unused inputs must be handled properly. For instance, SET and CLEAR must be connected so that they are not asserted (for a '74, tie them HIGH; for a 4013, they are grounded). Unused inputs that have no influence on the outputs can be left unconnected (e.g., inputs to unused gates), except in CMOS, where they should be grounded to prevent output-stage current (more on that in Chapter 9). A dual synchronizer is available as the 74120, although it has not been widely used.

MONOSTABLE MULTIVIBRATORS

The monostable multivibrator, or “one-shot” (emphasis on the word “one”), is a variation of the flip-flop (which is sometimes called a bistable multivibrator) in which the output of one of the gates is capacitively coupled to the input of the other gate. The result is that the circuit sits in one state. If it is forced to the other state by a momentary input pulse, it will return to the original state after a delay time determined by the capacitor value and the circuit parameters (input current, etc.). It is very useful (some would say *too* useful!) for generating pulses of selectable width and polarity. Making one-shots with gates and RCs is tricky, and it depends on the details of the gate’s input circuit, since, for instance, you wind up with voltage swings beyond the supply voltages. Rather than encourage bad habits by illustrating such circuits, we will just treat the one-shot as an available functional unit. In actual circuits it is best to use a packaged one-shot; you construct your own only if absolutely necessary, e.g., if you have a gate available and no room for an additional IC package (even then, maybe you shouldn’t).

8.20 One-shot characteristics

Inputs

One-shots are triggered by a rising or falling edge at the appropriate inputs. The only requirement on the triggering signal is that it have some minimum width, typically 25ns to 100ns. It can be shorter or longer than the output pulse. In general, several inputs are provided so that several signals can trigger the one-shot, some on positive edges and some on negative edges (remember, a negative edge means a HIGH-to-LOW transition, not a negative polarity). The extra inputs can also be used to inhibit triggering. Figure 8.64 shows four examples.

Each horizontal row of the table represents a valid input triggering transition. For example, the ’121 will trigger when one of the *A* inputs makes a HIGH-to-LOW transition, if the *B* input and the other *A* input are both HIGH. The ’4538 is a dual CMOS monostable with OR gating at the input; if only one input is used, the other must be disabled, as shown. The ’121 has three inputs, with a combination of OR and AND gating (and triggering), as shown. Its *B* input is a Schmitt trigger, more forgiving with slowly rising or noisy input signals. This monostable also includes a not-too-good internal timing resistor you can use instead of *R*, if you’re feeling lazy. The ’221 is a dual ’121; CMOS users can get only the dual version. The popular ’123 is a dual monostable with AND input gating; unused inputs must be enabled. Note particularly that it triggers when RESET is disabled if both trigger inputs are already asserted. This is not a universal property of monostables, and it may or may not be desirable in a given application (it’s usually not). The ’423 is the same as the ’123, but without this “feature.”

When drawing monostables in a circuit diagram, the input gating is usually omitted, saving space and creating a bit of confusion.

Retriggerability

Most monostables, e.g., the 4538, ’123, and ’423 mentioned earlier, will begin a new timing cycle if the input triggers again during the duration of the output pulse. They are known as retriggerable monostables. The output pulse will be longer than usual if they are retriggered during the pulse, finally terminating one pulse width after the last trigger. The ’121 and ’221 are nonretriggerable; they ignore input transitions during the output pulse. Most retriggerable one-shots can be connected as nonretriggerable one-shots.

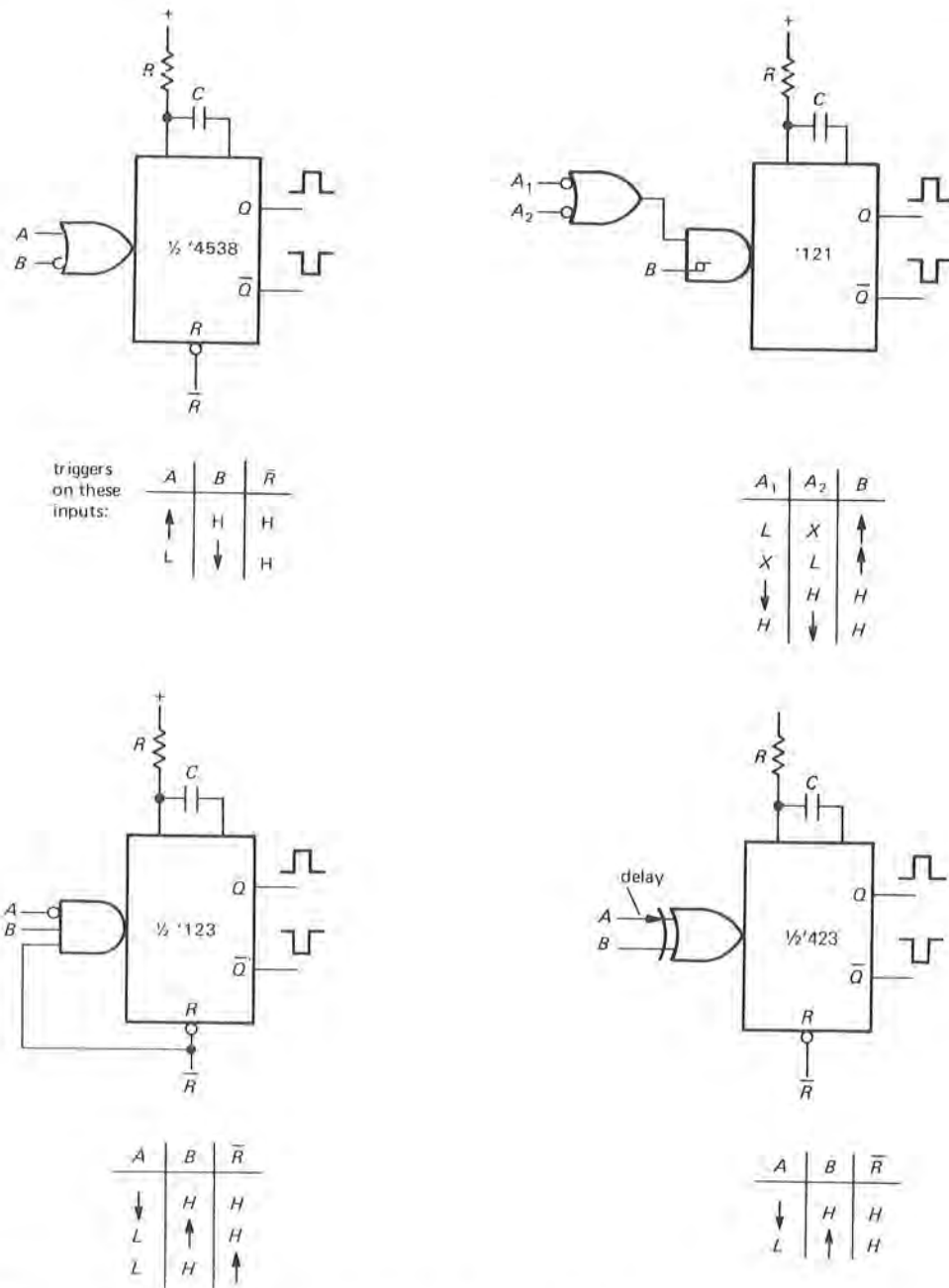


Figure 8.64. Four popular one-shots with their truth tables.

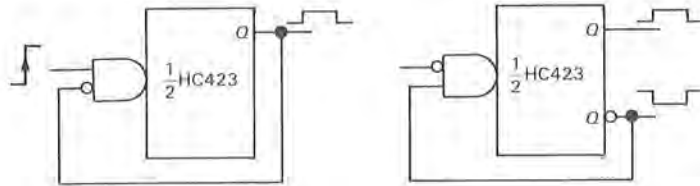


Figure 8.65. Nonretriggerable one-shot circuits.

Figure 8.65 shows an example that's easy to understand.

Resettability

Most monostables have a jam RESET input that overrides all other functions. A momentary input to the RESET terminal terminates the output pulse. The RESET input can be used to prevent a pulse during power-up of the logic system; however, see the preceding comment about the '123.

Pulse width

Pulse widths from 40ns up to milliseconds (or even seconds) are attainable with standard monostables, set by an external capacitor and (usually) resistor combination. A device like the 555 (Section 5.14) can be used to generate longer pulses, but its input properties are sometimes inconvenient. Very long delays are best generated digitally (see Section 8.23).

Table 8.8 at the end of the chapter lists most available monostables.

8.21 Monostable circuit example

Figure 8.66 shows a square-wave generator with independently settable rate and duty cycle (ratio of HIGH to LOW) and an input that permits an external signal to "hold" the output following a negative edge. Current mirror Q_1 – Q_3 generates a ramp at C_1 . When it reaches the threshold of the upper comparator at two-thirds V_+ , the one-shot is triggered and generates a $2\mu\text{s}$ positive pulse, putting n -channel VFET Q_4 into conduction and discharging the capacitor. C_1 therefore has a sawtooth

waveform going from ground to +8 volts, with rate set by potentiometer R_2 . The lower comparator generates an output square wave from the sawtooth, with duty cycle adjustable linearly between 1% and 99% via R_5 . Both comparators have a few millivolts of hysteresis (R_8 and R_9) to prevent noise-induced multiple transitions. The LM393 is a low-power dual comparator with input common-mode range right down to ground and open-collector outputs.

A feature of this circuit is its ability to synchronize (start/stop) to an externally applied control level. The HOLD input lets the driven circuit stop the oscillator at the next negative transition at the output. When HOLD is again brought LOW, the oscillator immediately resumes full cycles as if a falling edge had occurred at the time HOLD was released. The additional input to the 3-input NAND from the comparator output ensures that the circuit won't get stuck with C_1 charged up. In this circuit the one-shot pulse width has been chosen long enough to ensure that C_1 is fully discharged during the pulse.

8.22 Cautionary notes about monostables

Monostables have some problems you don't see in other digital circuits. In addition, there are some general principles involved in their use. First, a rundown on monostable pathology.

Some problems with monostables

Timing. One-shots involve a combination of linear and digital techniques. Since

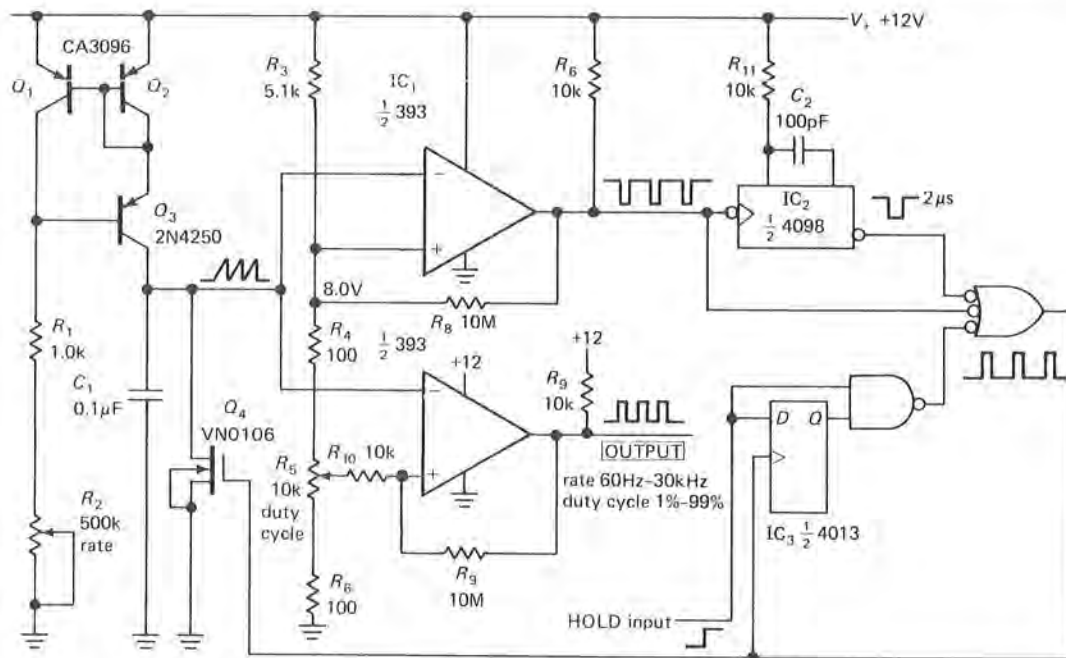


Figure 8.66. Autosynchronizing triggerable pulse generator.

the linear circuits have the usual problems of V_{BE} and h_{FE} variation with temperature, etc., one-shots tend to exhibit temperature and supply voltage sensitivity of output pulse width. A typical unit like the '4538 will show pulse-width variations of a few percent over a 0–50°C temperature range and over a $\pm 5\%$ supply voltage range. In addition, unit-to-unit variations give you a $\pm 10\%$ prediction accuracy for any given circuit. When looking at temperature and voltage sensitivity, it is important to remember that the chip may exhibit self-heating effects and that supply voltage variations *during the pulse* (e.g., small glitches on the V_+ line) may affect the pulse width seriously.

Long pulses. When generating long pulses, the capacitor value may be a few microfarads or more; in that case electrolytic capacitors are necessary. You have to worry about leakage current (which is insignificant with the smaller capacitor types), especially since most monostable

types apply voltage of both polarities across the capacitor during the pulse. It may be necessary to add a diode or transistor to prevent this problem, or to use a digital delay method instead (involving a clock and many cascaded flip-flop stages, as in Section 8.23). The use of an external diode or transistor will degrade temperature and voltage sensitivity and pulse-width predictability; it may also degrade retriggerable operation.

Duty cycle. With some one-shots the pulse width is shortened at high duty cycle. A typical example is the TTL 9600–9602 series, which has constant pulse width up to 60% duty cycle, decreasing about 5% at 100% duty cycle. The otherwise admirable '121 is considerably worse in this respect, with erratic behavior at high duty cycles.

Triggering. One-shots can produce substandard or jittery output pulses when triggered by too short an input pulse. There is

a minimum trigger pulse width specified, e.g., 50ns for the 'LS121, 140ns for the 4098 with +5 volt supply, and 40ns for the 4098 with +15 volt supply (CMOS is faster and has more output drive capability when operated at higher supply voltages).

Noise immunity. Because of the linear circuits in a monostable, the noise immunity is generally poorer than in other digital circuits. One-shots are particularly susceptible to capacitive coupling near the external R and C used to set the pulse width. In addition, some one-shots are prone to false triggering from glitches on the V_+ line or ground.

Specsmanship. Be aware that monostable performance (predictability of pulse width, temperature and voltage coefficients, etc.) may degrade considerably at the extremes of its pulse-width range. Specifications are usually given in the range of pulse widths where performance is good, which can be misleading. In addition, there can be a lot of difference from manufacturer to manufacturer in the performance of monostables of the same part number. Read the data sheets carefully!

Output isolation. Finally, as with any digital device containing flip-flops, outputs should be buffered (by a gate, inverter, or perhaps an interface component like a line driver) before going through cables or to devices external to the instrument. If a device like a one-shot drives a cable directly, the load capacitance and cable reflections may cause erratic operation to occur.

General considerations for using monostables

Be careful, when using one-shots to generate a train of pulses, that an extra pulse doesn't get generated at the "ends." That

is, make sure that the signals that enable the one-shot inputs don't themselves trigger a pulse. This is easy to do by looking carefully at the one-shot truth table, if you take the time.

Don't overuse one-shots. It is tempting to put them everywhere, with pulses running all over the place. Circuits with lots of one-shots are the mark of the neophyte designer. Besides the sort of problems just mentioned, you have the added complication that a circuit full of monostables doesn't allow much adjustment of the clock rate, since all the time delays are "tuned" to make things happen in the right order. In many cases there is a way to accomplish the same job without a one-shot, and that is to be preferred. Figure 8.67 shows an example.

The idea is to generate a pulse and then a second delayed pulse following the falling edge of an input signal. These might be used to set up and initiate operations that require that some previous operation be completed, as signaled by the input falling edge. Since the rest of the circuit is probably controlled by a "clock" square wave, let's assume that the signal at the D input falls synchronous with a clock rising edge. In the first circuit the input triggers the first one-shot, which then triggers the second one-shot at the end of its pulse.

The second circuit does the same thing with type D flip-flops, generating output pulses with width equal to one clock cycle. This is a synchronous circuit, as opposed to the asynchronous circuit using cascaded flip-flops. The use of synchronous methods is generally preferable from several standpoints, including noise immunity. If you wanted to generate shorter pulses, you could use the same kind of circuit, with the system clock divided down (via several toggling flip-flops) from a master clock of higher frequency. The master clock would then be used to clock the D flip-flops in this circuit. The use of several subdivided

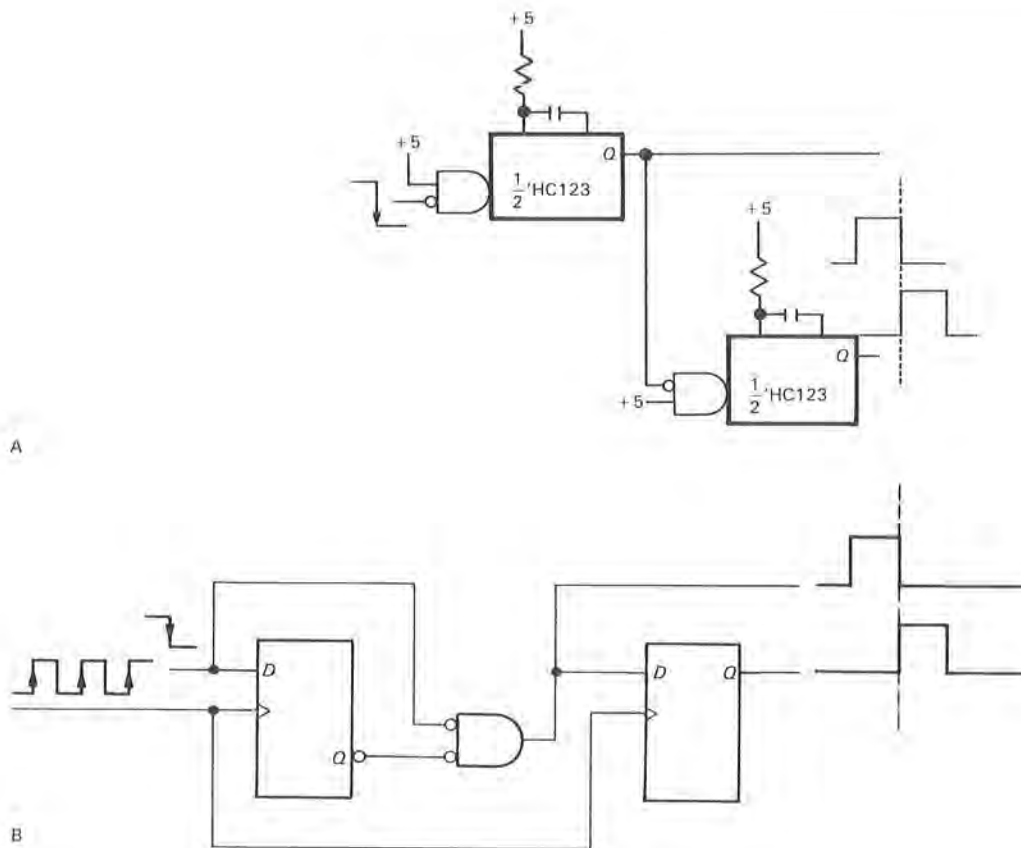


Figure 8.67. A digital delay can replace one-shot delays.

system clocks is common in synchronous circuits.

8.23 Timing with counters

As we have just emphasized, there are many good reasons for avoiding the use of monostables in logic design. Figure 8.68 shows another case where flip-flops and counters (cascaded toggling flip-flops) can be used in place of a monostable to generate a long output pulse. The '4060 is a 14-stage CMOS binary counter (14 cascaded flip-flops). A rising edge at the input brings Q HIGH, enabling the counter. After 2^{n-1} clock pulses, Q_n goes HIGH, clearing the flip-flop and the counter. This circuit generates an accurate long pulse whose length may be varied by factors of 2. The '4060

also includes internal oscillator circuitry that can substitute for the external clock reference. Our experience is that the internal oscillator has poor frequency tolerance and (in some HC versions) may malfunction.

You can get complete integrated circuits to implement timing with counters. The ICM7240/50/60 (Intersil, Maxim) have 8-bit or 2-digit internal counters and the necessary logic to make delays equal to an integral number of counts (1–255 or 1–99 counts); you can set the number either with “hardwired” connections or with external thumbwheel switches. The ICM7242 is similar, but with prewired divide-by-128 counter. Exar makes a close cousin, called the XR2243, which has a fixed divide-by-1024 counter.

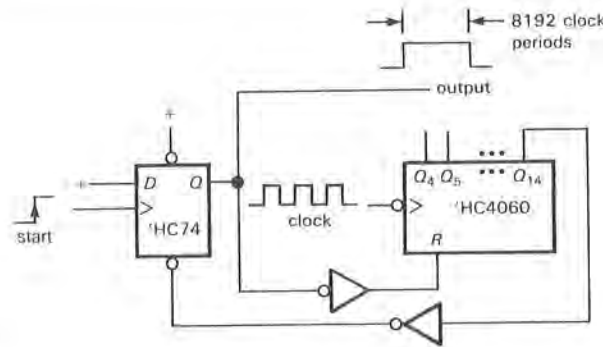


Figure 8.68. Digital generation of long pulses.

SEQUENTIAL FUNCTIONS AVAILABLE AS ICs

As with the combinational functions we described earlier, it is possible to integrate various combinations of flip-flops and gates onto a single chip. In the following sections we will present a survey of the most useful types, listed according to function.

As with pure combinational logic, *programmable* logic (PALs and GALs in particular) provides an attractive alternative to the use of prewired sequential functions. We'll talk about them, also, after looking at the standard functions.

8.24 Latches and registers

Latches and registers are used to "hold" a set of bits, even if the inputs change. A set of *D* flip-flops constitutes a register, but it has more inputs and outputs than necessary. Since you don't need separate clocks, or SET and CLEAR inputs, those lines can be tied together, requiring fewer pins and therefore allowing 8 flip-flops to fit in a 20-pin package. The popular '574 is an octal *D* register with positive clock edge and three-state outputs; the '273 is similar, but has a reset instead of three-state outputs. Figure 8.69 shows a quad *D* register with both true and complemented outputs.

The term "latch" is usually reserved for a special kind of register: one in which the outputs follow the inputs when

enabled, and hold the last value when disabled. Since the term "latch" has become ambiguous with use, the terms "transparent latch" and "type *D* register" are often used to distinguish these closely related devices. As an example, the '573 is the octal transparent-latch equivalent of the '574 *D* register.

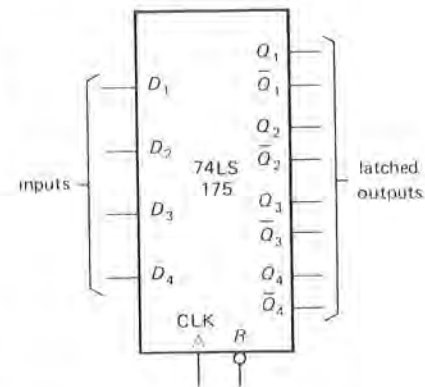


Figure 8.69. '175 4-bit *D* register.

Some variations on the latch/register are as follows: (a) random-access memories (RAMs), which let you write to, and read from, a (usually large) set of registers, but only one (or at most a few) at a time; RAMs come in sizes from a handful of bytes up to 1M bytes or more and are used primarily for memory in microprocessor systems (see Chapters 10 and 11); (b) addressable latches, a multibit latch that lets you update individual bits while keeping the others unchanged; (c) a latch or register built into a larger chip, for example a

digital-to-analog converter; such a device only needs the input applied momentarily (with appropriate clocking edge), since an internal register can hold the data.

Table 8.9 at the end of the chapter lists most of the useful registers and latches. Note features such as input enable, reset, three-state outputs, and “broadside” pinout (inputs on one side of the chip, outputs on the other); the latter is very convenient when you are laying out a printed-circuit board.

8.25 Counters

As we mentioned earlier, it is possible to make a “counter” by connecting flip-flops together. There is available an amazing variety of such devices as single chips. Here are some of the features to look for:

Size

You can get BCD (divide-by-10) and binary (or *hexadecimal*, divide-by-16) counters in the popular 4-bit category. There are larger counters, up to 24 bits (not all available as outputs), and there are *modulo-n* counters that divide by an integer n , specified as an input. You can always cascade counters (including synchronous types) to get more stages.

Clocking

An important distinction is whether the counter is a “ripple” counter or a “synchronous” counter. The latter clocks all flip-flops simultaneously, whereas in a ripple counter each stage is clocked by the output of the previous stage. Ripple counters generate transient states, since the earlier stages toggle slightly before the later ones. For instance, a ripple counter going from a count of 7 (0111) to 8 (1000) goes through the states 6, 4, and 0 along the way. This doesn’t cause trouble in well-designed circuits, but it would in a circuit that used

gates to look for a particular state (this is a good place to use something like a D flip-flop, so that the state is examined only at the clock edge). Ripple counters are slower than synchronous counters, because of the accumulated propagation delays. Ripple counters clock on negative-going edges for easy expandability (by connecting the Q output of one counter directly to the clock input of the next); synchronous counters clock on the positive edge.

We favor the ’160–’163 family of 4-bit synchronous counters for most applications that don’t require some special feature. The ’590 and ’592 are good 8-bit synchronous counters. Figure 8.70 shows the ’390 dual BCD ripple counter.

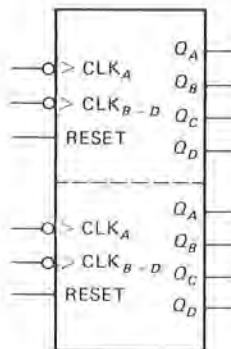


Figure 8.70. ’390 dual BCD ripple counter.

Up/down

Some counters can count in either direction, under control of some inputs. The two possibilities are (a) an U/D' input that sets the direction of count and (b) a pair of clocking inputs, one for UP, one for DOWN. Examples are the ’191 and ’193, respectively. The ’569 and ’579 are useful 8-bit up/down counters.

Load and clear

Most counters have data inputs so that they can be preset to a given count. This

is handy if you want to make a modulo- n counter, for example. The load function can be either synchronous or asynchronous: the '160-'163 have synchronous load, which means that data on the input lines are transferred to the counter coincident with the next clock edge, if the LOAD' line is also asserted LOW; the '190-'193 are asynchronous, or *jam-load*, which means that input data are transferred to the counter when LOAD' is asserted, independent of the clock. The term "parallel load" is sometimes used, since all bits are loaded at the same time.

The CLEAR (or RESET) function is a form of presetting. The majority of counters have a jam-type CLEAR function, though some have synchronous CLEAR; for example, the '160/161 are jam CLEAR, while the '162/163 are synchronous CLEAR.

Other counter features

Some counters feature latches on the output lines; these are always of the transparent type, so the counter can be used as if no latch were present. (Keep in mind that any counter with parallel-load inputs can function as latch, but you can't count at the same time as data are held, as you can with a counter/latch chip.) The combination of counter plus latch is sometimes very convenient, e.g., if you want to display or output the previous count while beginning a new counting cycle. In a frequency counter this would allow a stable display, with updating after each counting cycle, rather than a display that repeatedly gets reset to zero and then counts up.

There are counters with three-state outputs. These are great for applications where the digits (or 4-bit groups) are multiplexed onto a bus for display or transfer to some other device. An example is the '779, an 8-bit synchronous binary counter whose three-state outputs also serve as parallel inputs; by sharing input/output lines, the

counter fits in a 16-pin package. The '593 is similar, but in a 20-pin package.

If you want a counter to use with a display, there are several that combine counter, latch, 7-segment decoder, and driver on one chip. An example is the 74C925-74C928 series of 4-digit counters. Another amusing chip is the TIL306/7, a counter *with display* on one chip: You just look at the IC, which lights up with a digit telling the count! Figure 8.71 shows a nice LSI (large-scale-integration) counter circuit that doesn't require a lot of support circuits.

Table 8.10 at the end of the chapter lists most of the counter chips that you might want to use. Many of them are only available in one family (e.g., LS or F), so be sure to check the data books before you design with them.

8.26 Shift registers

If you connect a series of flip-flops so that each Q output drives the next D input, and all clock inputs are driven simultaneously, you get what's called a "shift register." At each clock pulse the pattern of 0's and 1's in the register shifts to the right, with the data at the first D input entering from the left. As with flip-flops, the data present at the serial input just prior to the clock pulse are entered, and there is the usual propagation delay to the outputs. Thus they may be cascaded without fear of a logic race. Shift registers are very useful for conversion of parallel data (n bits present simultaneously, on n separate lines) to serial data (one bit after another, on a single data line), and vice versa. They're also handy as memories, particularly if the data are always read and written in order. As with counters and latches, shift registers come in a pleasant variety of prefab styles. The important things to look for are the following:

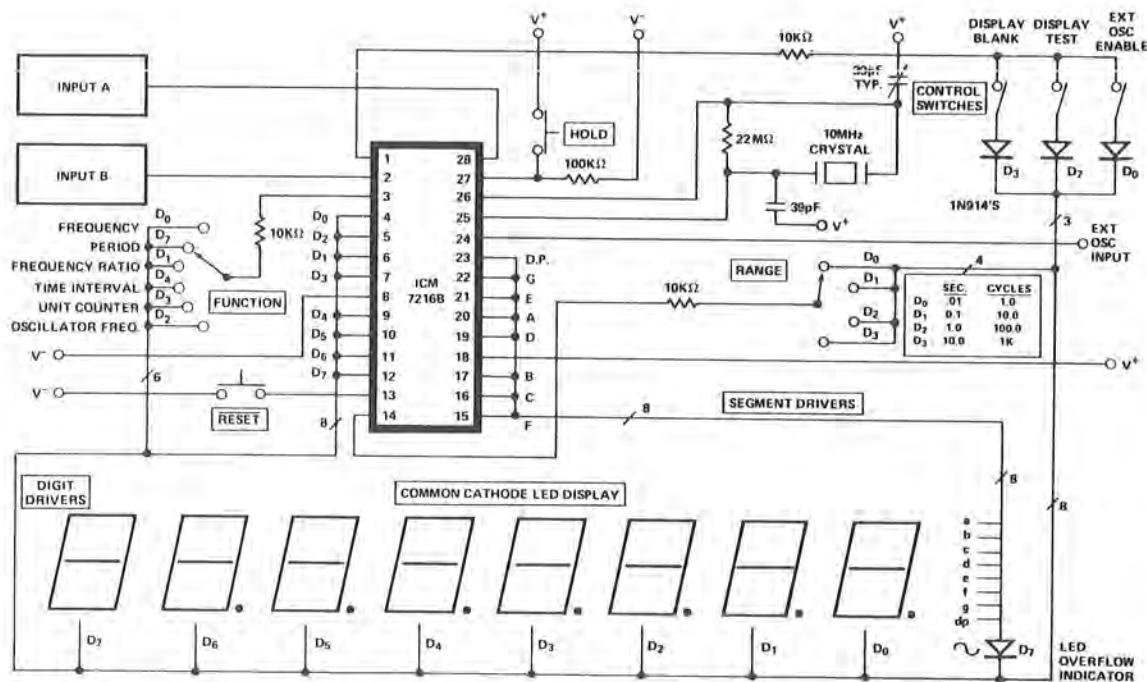


Figure 8.71. Intersil 7216 8-digit 10MHz universal counter on a chip. (Courtesy of Intersil, Inc.)

Size

The 4-bit and 8-bit registers are standard, with some larger sizes available (up to 64 bits or more). There are even variable-length registers (e.g., the 4557: 1 to 64 stages, set by a 6-bit input).

Organization

Shift registers are usually 1 bit wide, but there are also dual-, quad-, and hex-width registers. Most shift registers only shift right, but there are bidirectional registers like the '194 and '323 that have a "direction" input (Fig. 8.72). Watch out for trickery like the "bidirectional" '95, which can shift left only by tying each output bit to the previous input, then doing a parallel load.

Inputs and outputs

Small shift registers can provide parallel inputs or outputs, and usually do; an example is the '395, a 4-bit parallel-in, parallel-out (PI/PO) shift register with three-state outputs. Larger registers may only provide *serial* input or output, i.e., only the input to the first flip-flop or the output from the last is accessible. In some cases a few selected intermediate taps are provided. One way to provide both parallel input and output in a small package is to share input and output (three-state) on the same pins, e.g., the '299, an 8-bit bidirectional PI/PO register in a 20-pin package. Some shift registers include a latch at the input or output, so shifting can go on while data are being loaded or unloaded.

As with counters, parallel LOAD and

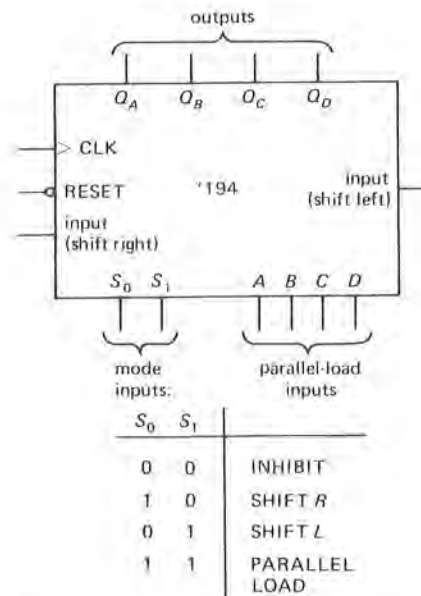


Figure 8.72. '194 4-bit bidirectional shift register.

CLEAR can be either synchronous or jam-load; for example, the '323 is the same as the '299, but with synchronous clear.

Table 8.11 at the end of the chapter lists the shift registers you're likely to use. As always, not all types are available in all logic families; be sure to check the data books.

RAMs as shift registers

A random-access memory can always be used as a shift register (but not vice versa) by using an external counter to generate successive addresses. Figure 8.73 shows the idea. An 8-bit synchronous up/down counter generates successive addresses for a 256-word $\times$ 4-bit CMOS RAM. The combination behaves like a quad 256-bit shift register, with left/right direction of shift selected by the counter's UP/DOWN' control line. The other inputs of the counter are shown enabled for counting. By choosing a fast counter and memory, we were able to achieve a maximum clocking rate

of 30MHz (see timing diagram), which is the same as that of an integrated (but much smaller) HC-type shift register. This technique can be used to produce very large shift registers, if desired.

EXERCISE 8.28

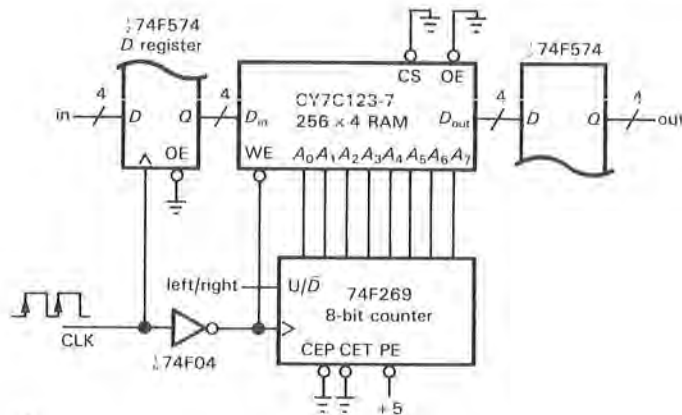
In the circuit of Figure 8.73, input data seem to go into the same location that output data are read from. Nevertheless, the circuit behaves identically to a classic 256-word shift register. Explain why.

8.27 Sequential PALs

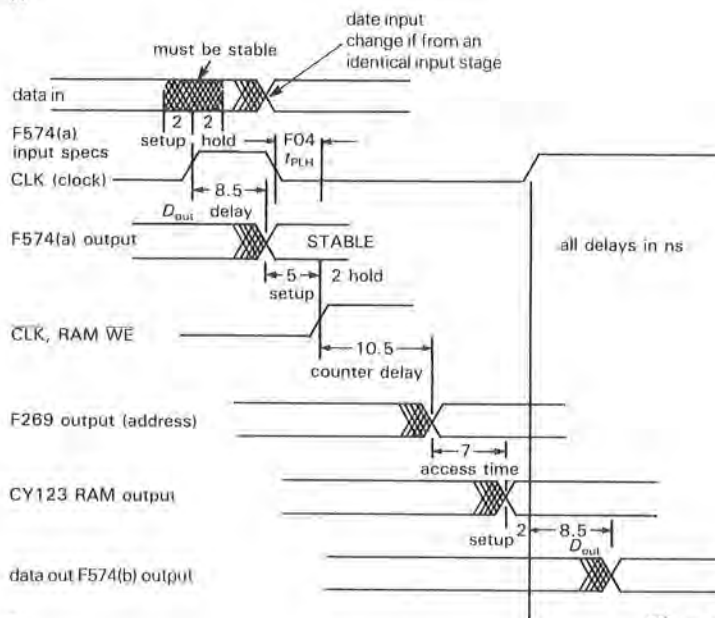
The combinational (gates-only) PALs we talked about in Section 8.15 belong to a larger family that includes devices with various numbers of on-chip *D*-type registers (called "registered PALs"). Typical of these PALs is the 16R8, shown in Figure 8.74. The programmable-AND/fixed-OR array typical of combinational PALs generates the input levels for 8 synchronously clocked *D*-type registers with three-state outputs; the register outputs (and their inverts) are available, along with the standard input pins, as inputs to the logic array. If you look back at Figure 8.57, you'll see that a registered PAL is a general-purpose sequential circuit element; within limits set by the number of registers and gates available, you can construct just about anything you want. For instance, you could make a shift register or counter, or some of both! In practice, you're more likely to make some custom piece of logic that is part of a larger circuit, for which the alternative is "discrete" logic built with gates and flip-flops. Let's look at some examples.

Hand-generated fuse maps

Simple designs can be implemented in PALs by figuring out the logic, then burning the appropriate pattern into the fuse array with a "PAL programmer." As an



A



B

F04 t_{PLH} delay	3.7ns	
F374 D_{out} delay	8.5ns	
CYC123 setup time	5	} 13.5 - 3.7 = 9.8ns HI
F269 counter out delay	10.5	
CYC123 access time	7	} 19.5 + 3.7 = 23.2ns LO
F374 D_{in} setup time	2	
shift clock period(min)	33 ns	

C

Figure 8.73. A. Large shift register made from RAM plus counter; the slash indicates multiple lines, in this case a 4-bit-wide data path.

B. Timing diagram to calculate maximum clocking rate, assuming worst-case timing specifications.

C. Calculation showing worst-case sequence of delays in a single clock cycle.

example, let's imagine we want a 4-input multiplexer with latched outputs. We can write the logic equation for the multiplexer portion (i.e., the D -input to the flip-flop) as

$$Q.d = I_0 * S'_0 * S'_1 + I_1 * S_0 * S'_1 \\ + I_2 * S'_0 * S_1 + I_3 * S_0 * S_1$$

where the inputs S_0 and S_1 address the selected input I_0 – I_3 and “*” and “+” represent AND and OR. A registered PAL makes it easy to latch the result. Note that we have used the OR of 3-input ANDs, rather than first decoding the select address in 2-input ANDs, because we

Logic Diagram

16R8

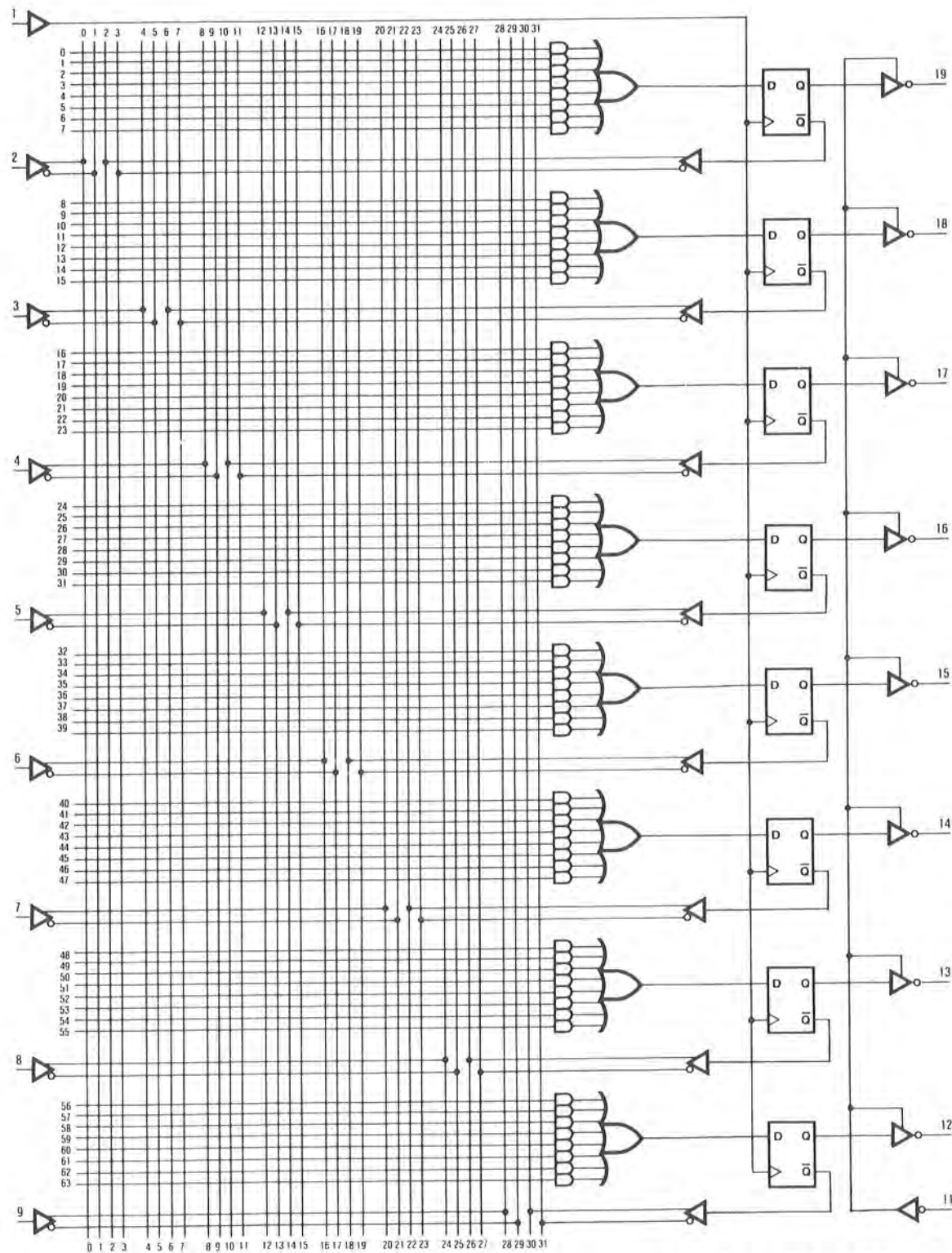


Figure 8.74. The 16R8 registered PAL has 8 external inputs, 8 outputs, a clock, and a three-state control line. The register outputs are also available as input to the AND array. (Diagram courtesy of Advanced Micro Devices of Sunnyvale, California.)

529

are constrained to use a sum of products (it's also faster). The finished design is shown in Figure 8.75. (Note: There's a subtlety to this circuit; see "Glitches" at the end of this section.)

PALASM

For circuits of any reasonable complexity, some PAL logic design aids are needed. For example, a 16L8 PAL has 2048 fuses; a complex design might require you to blow a few hundred of them, and unless you're unusually compulsive you're unlikely to succeed in manually specifying them all correctly to a PAL programmer.

PALASM (PAL Assembler; trademark of Monolithic Memories Inc.), developed by Monolithic Memories (the inventor of the PAL), was one of the first design aids. It takes Boolean expressions, like the one we wrote above, and converts them to fuse maps. No logic minimization is performed by the program, so you have to do most of the hard work yourself. However, PALASM does let you input a set of test states (called "test vectors"), giving you back the outputs that would result from your Boolean specification. In that way you can debug your equations before making the PAL.

PALASM is widely available. There are FORTRAN source code listings, and versions to run on popular microcomputers, which can then send the fuse map to a PAL programmer (in standardized

"JEDEC" format) via a serial port. Many PAL programmers with built-in microprocessors include a resident PALASM; some examples are Data I/O, Digilec, Stag, and Structured Design. With these you just hook up a terminal and you're in business.

ABEL and CUPL

PALASM is a help, but the serious PAL user needs more. High-level programmable-logic languages like ABEL (from Data I/O) and CUPL (from Logical Devices, Inc.) make programming PALs (and PLAs) easy work. They let you specify logic either by Boolean expressions or truth tables; for sequential circuits you give the states and transition rules. Like any good high-level language, you can define arrays (for a set of signals, e.g., an address bus), expressions, and intermediate values, then use them in later expressions.

These languages are smart enough to convert truth tables to logic equations, then minimize them (as well as any Boolean expressions you supplied) via logical identities, finally resulting in a form that fits the logic constraints of the device (e.g., sum of products, for a PAL). Instead of writing down the explicit logic expression for a range of values, you can just write something like ADDR:[10..FF], which will be converted to appropriate logic. These languages also let you specify test vectors, with which it tests your specified design;

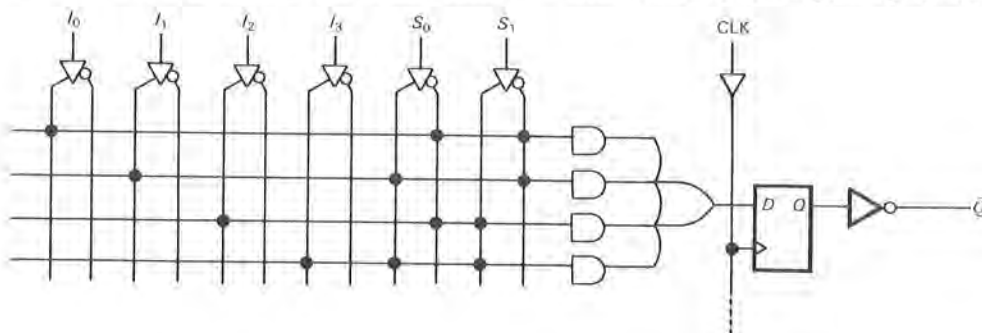


Figure 8.75. Latched 4-input multiplexer implemented in a PAL.

the test vectors can also be sent to the programmer to test the actual programmed chip. Finally, these languages provide standardized documentation of the finished chip, which is essential when you've got to debug a gadget with these mystery devices in it.

Just to make these ideas concrete, let's look at both combinational and sequential design examples using CUPL.

CUPL example: 7-segment-to-hex encoder (combinational). There are times when you might like to use an available LSI chip that performs a convenient function (e.g., a stopwatch or calculator) as part of an instrument you're building. The trouble is that these LSI chips usually provide outputs to drive 7-segment displays directly, rather than the hexadecimal (or BCD) outputs that you want. Therefore let's design an "encoder" chip that converts from 7-segment back to 4-bit binary, a function that is not available as a standard IC (although there is a 7-segment-to-BCD encoder, the 74C915).

The inputs are the individual segment signals, which are always labeled *a*–*g*; see Figure 8.76, which also shows how the digits A–F are represented with 7-segment displays. Note that it is possible to represent 9 and *C* in two ways, both of which should be correctly recognized by our logic. For the PAL, we choose a 16L8,

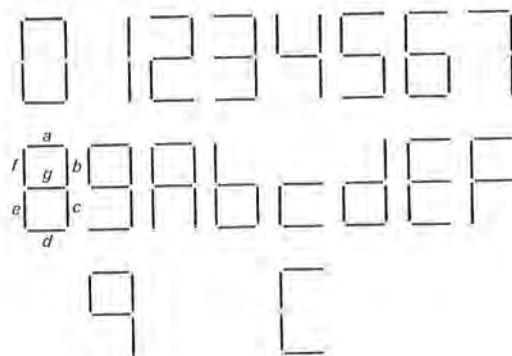


Figure 8.76. 7-segment display codes.

a 20-pin combinational part whose logic was shown in Figure 8.45.

Figure 8.77 is the input to CUPL. We assigned the (positive-true) segment drive signals *a*–*g* as inputs, and the (negative-true) hexadecimal bits D0–D3 as outputs. CUPL lets you define intermediate variables that can be used in later expressions; in this case it is convenient to define the obvious variables *zero* through *hexf*, the possible displayed digits in terms of the segment inputs. These are simply large product (AND) terms of the input segment variables, which you can read from the digit shapes in Figure 8.76. Finally, each binary output bit is written as the sum (OR) of the digit variables in which that bit is set. We've used negative-true levels because the 16L8 is an AND-NOR array, thus minimizing logic. This completes the logic specification to CUPL.

EXERCISE 8.29

Check for yourself that we did our work right by drawing some of the displayed characters as specified by our intermediate variables *zero* through *hexf*.

CUPL first uses the intermediate variable definitions to write the D0–D3 expressions directly in terms of the input variables *a*–*g*, a job that an assembler like PALASM would force us to do initially. At this point the logic equations are in the desirable AND-NOR form. However, we're not finished yet, because the 16L8 (and all other combinational PALs) permit at most 7 product terms in each sum, whereas we have 9, 8, 9, and 10, respectively, for the outputs D0–D3. One solution would be to string each output through a second OR gate, in order to get enough product terms in the sum; this is generally considered poor form because it doubles the propagation delay, though it wouldn't matter in a slow application like this. The better solution is to perform a logic minimization, using logic identities, DeMorgan's formula, etc.


```

/** Inputs **/

PIN 1  = a   ;    /* segment a */
PIN 2  = b   ;    /* segment b */
PIN 3  = c   ;    /* segment c */
PIN 4  = d   ;    /* segment d */
PIN 5  = e   ;    /* segment e */
PIN 6  = f   ;    /* segment f */
PIN 7  = g   ;    /* segment g */


/** Outputs **/

PIN 19 = !D3  ;    /* msb of hex encode */
PIN 18 = !D2  ;    /*
PIN 17 = !D1  ;    /*
PIN 16 = !D0  ;    /* lsb


/** Declarations and Intermediate Variable Definitions **/

zero = a & b & c & d & e & f & !g ;
one  = !a & b & c & !d & !e & !f & !g ;
two  = a & b & !c & d & e & !f & g ;
three = a & b & c & d & !e & !f & g ;
four  = !a & b & c & !d & !e & f & g ;
five  = a & !b & c & d & !e & f & g ;
six   = a & !b & c & d & e & f & g ;
seven = a & b & c & !d & !e & !f & !g ;
eight = a & b & c & d & e & f & g ;
nine  = a & b & c & !d & !e & f & g
      # a & b & c & d & !e & f & g ;    /* two ways */
hexa  = a & b & c & !d & e & f & g ;
hexb  = !a & !b & c & d & e & f & g ;
hexc  = !a & !b & !c & d & e & !f & g
      # a & !b & !c & d & e & f & !g ;    /* two ways */
hexd  = !a & b & c & d & e & !f & g ;
hexe  = a & !b & !c & d & e & f & g ;
hexf  = a & !b & !c & !d & e & f & g ;


/** Logic Equations **/

D3 = eight # nine # hexa # hexb # hexc # hexd # hexe # hexf ;
D2 = four # five # six # seven # hexc # hexd # hexe # hexf ;
D1 = two # three # six # seven # hexa # hexb # hexe # hexf ;
D0 = one # three # five # seven # nine # hexb # hexd # hexf ;

```

Figure 8.77. 7-segment-to-hex encoder specification, written in CUPL language.

We ran CUPL's minimizer, producing the product terms shown in Figure 8.78. Luckily, all fit within the 7-product constraint. CUPL also draws a fuse map for you (Fig. 8.79). Of course, you don't pro-

gram the PAL from that, but use instead a direct download of the universal JEDEC programming format. In this example, CUPL has obviously made a laborious problem simple.

** Expanded Product Terms **

```

D0 =>
  a & b & c & d & !e & g
# a & b & c & !e & f & g
# a & c & d & !e & f & g
# b & c & !d & !e & !f & !g
# !a & !b & c & d & e & f & g
# a & !b & !c & !d & e & f & g
# !a & b & c & d & e & !f & g

D1 =>
  a & !b & !c & e & f & g
# !b & c & d & e & f & g
# a & b & c & !d & e & f & g
# a & b & !c & d & e & !f & g
# a & b & c & d & !e & !f & g
# a & b & c & !d & !e & !f & !g

D2 =>
  a & !b & !c & d & e & f
# a & !b & c & d & f & g
# a & !b & !c & e & f & g
# !a & b & c & !d & !e & f & g
# !a & b & c & d & e & !f & g
# !a & !b & !c & d & e & !f & g
# a & b & c & !d & !e & !f & !g

D3 =>
  a & b & c & f & g
# a & !b & !c & d & e & f
# a & !b & !c & e & f & g
# !a & !b & c & d & e & f & g
# !a & b & c & d & e & !f & g
# !a & !b & !c & d & e & !f & g

D0.oe => 1
D1.oe => 1
D2.oe => 1
D3.oe => 1

```

Figure 8.78. 7-segment-to-hex encoder: minimized product terms.

CUPL example: vending machine (sequential). An arbitrary state machine (Section 8.18) has a set of *states* and a set of *transition rules* for moving between those states at each clock edge. In general, the transition rules depend both on the present state and on the particular combination of input levels present at the next clock edge. You can implement a state machine in programmable logic containing registers if (a) there are enough register bits to represent all possible states (e.g., with 4 registers you could have up to 16 states) and

(b) there are enough inputs and logic gates to implement the transition rules.

As an example, let's design a registered PAL for the state diagram of Figure 8.80. It's a vending machine, and it is supposed to disgorge a bottle of fizzy sweet liquid when 25 cents (or more) has been deposited. There is some sort of coin interface that gobbles up and recognizes money and sends to our PAL a 2-bit input (C1, C0), valid for one clock edge, indicating the coin just deposited (01 = nickel, 10 = dime, 11 = quarter, 00 = slug or no coin).

```

** Fuse Plot **

Pin #19
0000 -----
0032 x-x-x-----x--x-----
0064 -xx--x--x--x--x--x-----
0096 -xx--x--x--x--x--x-----
0128 -x-xx--x--x--x--x--x-----
0160 x--xx--x--x--x--x--x-----
0192 -x-x-x--x--x--x--x--x-----
0224 xxxxxxxxxxxxxxxxxxxxxxxxxxxx

Pin #18
0256 -----
0288 -xx--x--x--x--x--x-----
0320 -xx-x--x--x--x--x-----
0352 -xx--x--x--x--x--x-----
0384 x--xx--x--x--x--x-----
0416 x--xx--x--x--x--x-----
0448 -x-x-x--x--x--x--x-----
0480 x-x-x--x--x--x--x-----

Pin #17
0512 -----
0544 -xx--x--x--x--x--x-----
0576 -x--x--x--x--x--x-----
0608 x-x-x--x--x--x--x-----
0640 x-x--x--x--x--x--x-----
0672 x-x-x--x--x--x--x-----
0704 x-x-x--x--x--x--x-----
0736 xxxxxxxxxxxxxxxxxxxxxxxxxxxx

Pin #16
0768 -----
0800 x-x-x--x--x--x-----
0832 x-x-x--x--x--x-----
0864 --x-x--x--x--x--x-----
0896 x--x--x--x--x--x-----
0928 -x-xx--x--x--x--x-----
0960 -xx--x--x--x--x-----
0992 x--xx--x--x--x--x-----

```

LEGEND X : fuse not blown
 - : fuse blown

Figure 8.79. 7-segment-to-hex encoder; fuse map.

The state machine's job is to add up the total deposited and generate an output called *bottle* when there's enough money.

Figure 8.81 shows the specification, in CUPL's state-machine syntax. As before, we begin by defining input and output pins. Note that we've added a *reset* input so that you can initialize to the state S0 (no money). Next we define the states, then the rules for moving between them. If any outputs, either registered

or combinational, need to be generated during states or transitions between states, they are specified at the same time. In this example, for instance, the output *bottle* has been specified as a separate output register, so that no output state decoding is needed. In fact, this is the only output needed, and the state-machine bits Q0–2 could be implemented in internal registers that don't generate outputs directly; some programmable logic devices have such

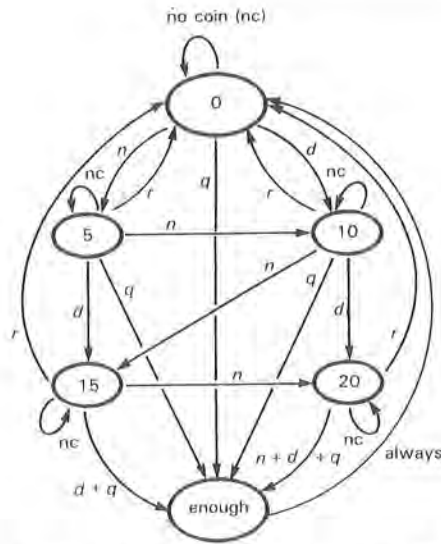


Figure 8.80. Vending machine state diagram.

“buried” registers, in addition to the usual output registers.

Note that you have to specify explicitly the transition from a state to itself, as we have done here for the input *no coin*. An unspecified condition implicitly resets the state to all zeros. That is because these conditionals are compiled into combinational logic to assert the *D* inputs of the registers, and thus if the condition is not met, the corresponding *D* input is not asserted.

Figure 8.82 shows the output from CUPL. There is nothing obvious or simple about the logic, because both the machine state (S0–S5) and the input variable (C0–1) are specified as binary numbers, whereas the logic operates on individual

```

/** Inputs **/

Pin 1 = clk      ; /* clock -- positive edge */
Pin 3 = c0       ; /* coin type -- low bit   */
Pin 4 = c1       ; /* coin type -- high bit  */
Pin 6 = reset    ; /* reset input          */

/** Outputs **/

Pin 18 = !Q0     ; /* bit 0 of state variable */
Pin 17 = !Q1     ; /* bit 1                    */
Pin 16 = !Q2     ; /* bit 2                    */
Pin 15 = !bottle ; /* bottle disgorge command */

/* Define machine states with symbolic names;
   "enough" = 25 cents or more */

#define S0      'b'000
#define S5      'b'001
#define S10     'b'010
#define S15     'b'011
#define S20     'b'100
#define ENOUGH  'b'101

/* define intermediate variables */

nocoin = !c0 & !c1 & !reset;
nickel = c0 & !c1 & !reset;
dime   = !c0 & c1 & !reset;
quarter = c0 & c1 & !reset;

/* Define state bit variable field */

field statebit = [Q2..0] ;

```

Figure 8.81. Vending machine specification (CUPL). (Continued on next page.)

```

/* Transition rules for vending machine */
sequence statebit {
  present S0    if nocoin    next S0;
                if nickel    next S5;
                if dime      next S10;
                if quarter   next ENOUGH    out bottle;

  present S5    if nocoin    next S5;
                if nickel    next S10;
                if dime      next S15;
                if quarter   next ENOUGH    out bottle;

  present S10   if nocoin    next S10;
                if nickel    next S15;
                if dime      next S20;
                if quarter   next ENOUGH    out bottle;

  present S15   if nocoin    next S15;
                if nickel    next S20;
                if dime      next ENOUGH    out bottle;
                if quarter   next ENOUGH    out bottle;

  present S20   if nocoin    next S20;
                if nickel    next ENOUGH    out bottle;
                if dime      next ENOUGH    out bottle;
                if quarter   next ENOUGH    out bottle;

  present ENOUGH      next S0;  }

```

Figure 8.81. (cont.)

bits. Thus, the resulting logic does not bear much relation to the original state description (Fig. 8.81). In fact, the particular choice of states (ascending binary, 0–5) could have been chosen differently, completely changing the resulting logic. In this case, this example fits easily within the constraints of the 16R6 PAL (8 product terms per register); if it had not, we could have tried redefining the states, which often leads to a simpler logic realization. Note that the *reset* input acts by an overriding disassertion of all *D* inputs, which we forced by our definition of the intermediate variables *nocoin*, *nickel*, etc.

EXERCISE 8.30

Verify that the finished logic equations are correct, by checking a few transition rules. You might try all the transitions from 00, or a nickel or dime from some other state.

EXERCISE 8.31

Good vending machines give change. Redraw the state diagram (Fig. 8.80) so that there are states (how many?) for each possible amount of change. Modify the transition rules accordingly. Make sure your modified vending machine still does its primary job – dispensing bottles.

EXERCISE 8.32

Draw a state diagram and transition rules for an electronic combination lock: It should open only if four digits are entered in the correct order. Any error should reset it.

Hints for using programmable logic devices (PLDs)

For certain jobs, PLDs really hit the spot. Here are the most important uses and advantages of PLDs:


```

** Expanded Product Terms **

Q0.d =>
    !Q0 & !Q1 & c0 & !reset
    # !Q0 & !Q2 & c0 & !reset
    # Q0 & !Q2 & !c0 & !reset
    # Q0 & !Q2 & c1 & !reset
    # !Q0 & !Q1 & Q2 & c1 & !reset

Q1.d =>
    !Q1 & !Q2 & !c0 & c1 & !reset
    # !Q0 & Q1 & !Q2 & !c1 & !reset
    # Q1 & !Q2 & !c0 & !c1 & !reset
    # Q0 & !Q1 & !Q2 & c0 & !c1 & !reset

Q2.d =>
    !Q0 & !Q1 & Q2 & !reset
    # Q1 & !Q2 & c1 & !reset
    # !Q2 & c0 & c1 & !reset
    # Q0 & Q1 & !Q2 & c0 & !reset

bottle.d =>
    !Q2 & c0 & c1 & !reset
    # !Q0 & !Q1 & Q2 & c0 & !reset
    # !Q0 & !Q1 & Q2 & c1 & !reset
    # Q0 & Q1 & !Q2 & c1 & !reset

statebit =>
    Q2 , Q1 , Q0

```

Figure 8.82. Vending machine: CUPL output.

State machines. As in the foregoing example, the programmable logic device is a natural for an arbitrary synchronous state machine. You would be foolish to use an array of *D* flip-flops and discrete combinational logic when a PAL does the job in one inexpensive and powerful package.

Replacing "random" logic. Within many circuits you find little knots and tangles of gates, inverters, and flip-flops, scornfully referred to as *random logic*, or *glue*. A PLD will generally cut the package count by a factor of 4 or more. It also generally results in cleaner design, because the enormous number of gates available means that you can do all your gating at the *inputs* to the registers (resulting in strictly synchronous outputs), instead of the gate-conserving method of also combining register outputs via gates. With the latter the outputs are not strictly synchronous; Figure 8.83.

Flexibility. Sometimes you're not quite sure how you ultimately want some circuit to work, yet you must finish the design so that you can play around with it. PLDs are great here, because you can substitute one with different programming at some later stage, without the rewiring you'd have to do if you had used discrete logic. With PLDs, the *circuit* is a form of software!

Multiple versions. PLDs make it possible to design a single circuit, then produce several different versions of the instrument by populating the board with different PLDs. For example, you could have a computer that could accept either 256K or 1M memory chips, with just a change of PAL.

Speed and inventory. With PLDs you can generally get the design job done more quickly (once you've learned the ropes). Furthermore, you only need to stock a few PLD types, rather than dozens of