

15

TIMING CIRCUITS

In this chapter we consider two timing circuits, the *monostable multivibrator* and the *astable multivibrator* and, also a general purpose *timer* capable of performing either operation as well as a wide variety of other timing functions.

The flip-flop circuit (also called a bistable multivibrator), it will be recalled, has two stable states, in either one of which it may remain permanently. The monostable multivibrator has only one permanently stable state and one quasi-stable state. In the monostable configuration, a triggering signal is required to induce a transition from the stable state to the quasi-stable state. The circuit may remain in its quasi-stable state for a time which is very long in comparison with the time of transition between states. Eventually, however, it will return from the quasi-stable state to its stable state, no external signal being required to induce this reverse transition.

Since, when it is triggered, the monostable circuit returns by itself to its original state after a time T , it is also known as a *one-shot*. Since it generates a rectangular waveform which can be used to gate other circuits, it is also called a *gating circuit*. Furthermore, since it generates a fast transition at a predetermined time T after the input trigger, it is also referred to as a *delay circuit*.

The astable multivibrator has two states, both of which are quasi-stable.

TIMING CIRCUITS 545

Without the aid of an external triggering signal the astable configuration will make successive transitions from one quasi-stable state to the other. Thus the astable circuit is an *oscillator* and is used as a generator of square waves or a clocking waveform.

15.1 CMOS MULTIVIBRATORS

CMOS gates are conveniently adaptable for use in monostable and astable multivibrators. In this section we consider monostable circuits using such gates.

We discussed CMOS gates in Chap. 8. The diagram of a two-input NOR gate is given in Fig. 8.9-2. This NOR gate is shown again in Fig. 15.1-1. Note the presence of the input protective diodes (Sec. 1.16). These diodes are bridged between each input and the power-supply terminals and are used (operating in conjunction with the resistive impedance of the input driving sources) to restrain the voltages on the gates of the MOS devices to the voltage range nominally from ground to V_{SS} . This restraint is necessary to assure that the gate voltages will never exceed the breakdown voltage of the oxide layer which insulates the gate from the semiconductor. Actually, the breakdown voltage may be of the order of 100 V. Hence, in ordinary operation (using supply voltages V_{SS} in the range of about 10 V) it is not likely that the breakdown voltage would be exceeded. There is, however, a special difficulty which is encountered with MOS devices generally and which results from the extremely high insulation resistance ($\approx 10^{12} \Omega$) of the gate. As a result, when no dc path is provided to allow charge leakage from a gate, even a static charge inadvertently established on a gate (as by touching an isolated input-gate terminal) may cause breakdown. Such breakdown is precluded by the presence of the input diodes.

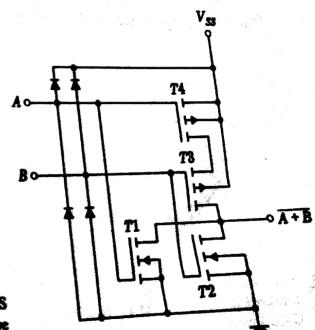
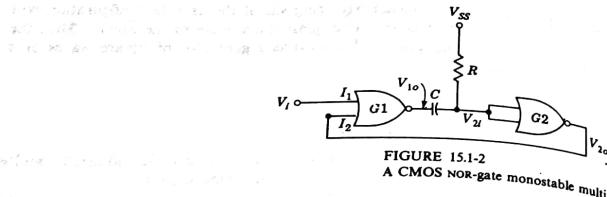


FIGURE 15.1-1
Schematic diagrams of two-input CMOS NOR gates, illustrating the input protective diodes (Motorola MC 14001).



When these gates are used in straightforward gate circuits, the input diodes may actually never be called upon to conduct. On the other hand, in multi circuits some of the diodes are essential, and if they were not already incorporated in the gate itself, it would be necessary to add them externally.

A basic CMOS monostable multi using two NOR gates is shown in Fig. 15.1-2. As indicated, the resistor R is returned to the supply voltage V_{ss} , which powers the gates. The gate $G2$ has its inputs joined since this gate is being used simply as an inverter.

Some details of the waveforms associated with the CMOS multi depend on the precise form of the input-output characteristic of the CMOS gate. A typical characteristic is shown in Fig. 8.4-1b. However, the monostable multi is used essentially to establish a time interval. Matters relating to this time interval, its range, stability, etc., are of interest while the exact waveforms are of no serious consequence. Accordingly, to simplify the discussion we shall consider that the input-output characteristic of the CMOS may be represented as in Fig. 15.1-3. Here we assume an abrupt change in output between V_{ss} to 0 when the input is at the transition voltage V_T . With this assumption, the waveforms of the monostable multi appear as in Fig. 15.1-4.

In the stable state the input V_{2i} to gate $G2$ will be at logic level 1; that is, $V_{2i} = V_{ss}$ (since MOS gates and hence $G2$ draw negligible input current).

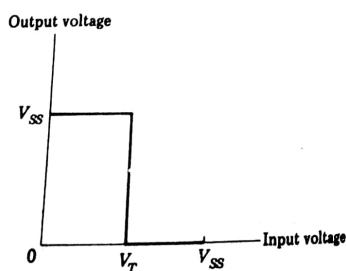
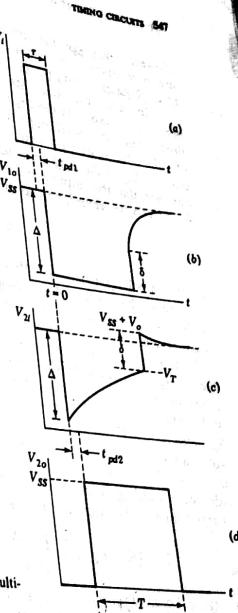


FIGURE 15.1-4
Waveforms of the CMOS astable multi-vibrator.

The output V_{2o} will be approximately at 0 V, that is, at logic level 0. The trigger input to initiate the quasi-stable state is applied at the input terminal I_1 of gate $G1$. With the input V_t and the input V_{2i} at I_2 both at logic 0, the output V_{1o} of gate $G1$ will be at logic 1, that is, at V_{ss} . Thus capacitor C is initially discharged.

Now let a short positive pulse be applied at I_1 , as indicated in Fig. 15.1-4a, of amplitude adequate to carry I_1 to logic level 1. Then after a propagation delay t_{pd1} due to $G1$, V_{1o} will drop abruptly (Fig. 15.1-4b), and this abrupt drop will be transmitted through C to the input of gate $G2$. Assuming that the drop is great enough in magnitude to carry the voltage V_{2i} (Fig. 15.1-4c) to logic level 0, the output V_{2o} (Fig. 15.1-4d) will rise to logic 1 after a propagation delay t_{pd2} . The gate input I_2 will then be held at logic 1, and the output of $G1$ will thereafter not be affected by the termination of the input trigger pulse V_t .



However, the trigger has served to initiate the quasi-stable state. Note that the minimum duration of the pulse V_t must exceed the sum of the propagation delays through gates G1 and G2, that is, $\tau = t_{p1} + t_{p2}$.

Initially, in the stable state, the current through R is zero and the voltage across C is zero. At the beginning of the quasi-stable state, when V_{1o} and V_{2i} drop, a current flows through R, continues through C, and sinks into the output of gate G1. If gate G1 were not being called upon to sink this current, the output V_{1o} would drop from its initial level V_{ss} down to ground. Suppose, however, that the output resistance of gate G1 is R_{o1} ; then the drop Δ in V_{1o} and V_{2i} , as can be verified, will not be equal to V_{ss} but will instead be given by

$$\Delta = \frac{R}{R + R_{o1}} V_{ss} \quad (15.1-1)$$

so that Δ will approach V_{ss} only if $R \gg R_{o1}$.

As C charges from V_{ss} through R and R_{o1} , the charging current decreases with time. Consequently the drop across R_{o1} decays as well, and during the quasi-stable interval T the waveform V_{1o} displays a downward tilt.

Starting at $t = 0$ (taken to be the time when V_{1o} and V_{2i} drop by amount Δ), the voltage V_{2i} rises from $V_{ss} - \Delta$ toward V_{ss} as an asymptotic limit. The time constant associated with this exponential waveform is $(R + R_{o1})C$. The quasi-stable state is terminated when V_{2i} reaches the transition voltage V_T . When V_{2i} passes V_T , V_{2o} returns to its initial value of 0 V and, as soon as the capacitor C has discharged the charge it has acquired, both V_{1o} and V_{2i} return as well to their initial values V_{ss} . The circuit will remain in this stable state indefinitely until again induced to make a transition to the quasi-stable state by an input pulse. The time of the quasi-stable state is (Prob. 15.1-1) C

$$T = (R + R_{o1})C \ln \left(\frac{\Delta}{V_{ss} - V_T} \right) \approx (R_o + R_{o1})C \ln \frac{V_{ss}}{V_{ss} - V_T} \quad (15.1-2)$$

In a typical case, with $V_T = V_{ss}/2$, we find $T = 0.7(R_o + R_{o1})C$.

The circuit which determines the charging and discharging of the timing capacitor C is shown in Fig. 15.1-5. Diode D represents the parallel combination of the diodes which connect the paralleled inputs of gate G2 to V_{ss} . Gate G1, as viewed from its output, is represented by the two switches S and S' and the "elements" R_{o1} and R' . In the stable state S' is closed, S is open, and the capacitor voltage V_C is $V_C = 0$ V. The quasi-stable state begins when S closes and S' opens and C charges from V_{ss} through R and R_{o1} . If R is large enough, then the voltage across the transistors (T1 and T2 in Fig. 15.1-1) will initially be low enough (and remain so as C charges) so that the transistors will operate in the triode region where they may reasonably be represented as a resistor R_{o1} .

The quasi-stable state ends when S opens and S' closes. During the quasi-stable state, the capacitor may acquire a voltage comparable to V_{ss} . At the end of the quasi-stable state C discharges through the diode D and through the "element" R' . The element R' is intended to represent the volt-ampere characteristic of the series combination of T3 and T4 in Fig. 15.1-1. Initially,

TIMING CIRCUITS 55

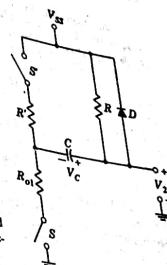


FIGURE 15.1-5
Circuit which determines the charging and discharging cycles of the timing capacitor C.

when V_C is large, the transistors may be in the saturation region so that the capacitor will discharge at a nominally constant rate. As V_C decreases, the transistors may be in the triode region so that the capacitor discharge will continue exponentially. Finally, after the voltage V_C drops to the point where the diode voltage falls below its cut-in voltage V_D (≈ 0.65 V), the last part of the capacitor discharge will take place through R' and R.

When S opens, S' closes, and diode D starts to conduct, the voltage V_{2i} will become $V_{ss} + V_D$ (at this point, since the diode may well be carrying a substantial current, we may reasonably take V_D to be about 0.75 V). Hence, as is to be seen in Fig. 15.1-4c, at the end of the quasi-stable state V_{2i} jumps abruptly from V_T to $V_{ss} + V_D$. The size of the jump is $\delta = V_{ss} + V_D - V_T$. Since the output of G1 is coupled to the input of G2 by the capacitor, and since the voltage across the capacitor itself cannot change abruptly, there is an equal jump δ in the waveform of V_{1o} . Thereafter V_{2i} and V_{1o} decay to their steady-state value V_{ss} . As noted, R' is not a simple resistor and D does not conduct to the very end. Nonetheless, for simplicity, in Fig. 15.1-4b and c we have indicated the discharge of C as generating a simple exponential waveform.

EXAMPLE 15.1-1 In the monostable multivibrator circuit of Fig. 15.1-2, $V_{ss} = 10$ V, $V_T = 5$ V, $C = 0.01 \mu\text{F}$, $R = 10 \text{ k}\Omega$, $R_{o1} = 500 \Omega$, and assume that R' is a "resistor" and let $R' = 1 \text{ k}\Omega$.

- (a) Find Δ , δ , and the voltage V_{1o} at the end of the quasi-stable state.
- (b) Find the time T of the quasi-stable state.
- (c) Estimate how long a time, after the end of the quasi-stable state, will be required for the capacitor to discharge to 0.1 V.
- (d) Suppose a second triggering pulse is applied at the time $V_C = 0.1$ V. How will the second timing interval compare with the first?

SOLUTION (a) From Eq. (15.1-1), $\Delta = [10/(10 + 0.5)](10) = 9.5 \text{ V}$; $\delta = V_{ss} + V_o$.
 $V_T = 10 + 0.75 - 5 = 5.7 \text{ V}$; when $V_{oi} = V_T = 5 \text{ V}$, the current through R and through
 R_{o1} is $(V_{ss} - V_T)/R$. The drop across R_{o1} is $R_{o1}(V_{ss} - V_T)/R = 0.5(5)/10 = 0.25 \text{ V} = V_{io}$.

(b) From Eq. (15.1-2),

$$T = (R + R_{o1})C \ln \left(\frac{V_{ss}}{V_{ss} - V_T} \right)$$

$$= (10 + 0.5) \times 10^3 \times 0.01 \times 10^{-6} \ln \left(\frac{10}{5} \right)$$

$$= 72 \mu\text{s}$$

(c) Assume that the diode maintains across itself a voltage 0.7 V (a compromise between 0.65 and 0.75) as long as the diode current is 0.1 mA or larger. Let $t = 0$ at the beginning of the capacitor discharge at which time the capacitor voltage is $V_C = 5 - 0.25 = 4.75 \text{ V}$. Then the capacitor discharge current is

$$I_C = \frac{4.75}{R'} e^{-t/R'C} = 4.75 e^{-10t} \text{ mA}$$

We find $I_C = 0.1 \text{ mA}$ at $t = 39 \mu\text{s}$. Just before the diode turns off the capacitor voltage V_C is $0.7 \text{ V} + I_C R' = 0.7 + 0.1 = 0.8 \text{ V}$. The decay from 0.8 to 0.1 V occurs as C discharges through the series combination of R and R' . This time is calculated from

$$V_C = 0.8 e^{-t/R + R'C} = 0.8 e^{-1.1 \times 10^{-4} t}$$

We find that $V_C = 0.1 \text{ V}$ when $t = 230 \mu\text{s}$. The total time is $39 + 230 = 269 \mu\text{s}$. Observe that the time required, after the quasi-stable state, for the circuit to recover very nearly to its initial situation is rather long in comparison with the duration of the quasi-stable state itself.

(d) As calculated in (b) the capacitor charges from 0 to 5 V in 72 μs . Let us assume for simplicity that the capacitor charges at a constant rate. This is a rough approximation but, for the accuracy desired, it is sufficient. Hence to charge from 0.1 to 5 V will require a time $(4.9/5)72 = 70.6 \mu\text{s}$. The change in timing is 1.4 μs and the percentage change is about 2 percent.

In Fig. 15.1-3 we have suggested that the transition between logic levels in the CMOS gate occurs at an input voltage which is about one-half the supply voltage V_{ss} . In practice, the transition region of a CMOS gate may vary from sample to sample and may occur anywhere in the range from about 30 percent V_{ss} to 70 percent V_{ss} . Hence, if in the circuit of Fig. 15.1-2 the R and C are set for a particular time of the quasi-stable state, this time may vary very considerably from circuit to circuit with different samples of the gate [see Eq. (15.1-2)]. On the other hand, the input-output characteristic of a CMOS gate displays very little sensitivity to temperature (see Fig. 1.15-3), so that the timing with any particular CMOS sample will be rather temperature-insensitive.

A number of circuits are available^{1,2} which considerably reduce the dependence of timing on CMOS samples. These circuits depend on the fact that while there may be considerable variability from sample to sample as we go from one integrated-circuit chip to another, samples of gates fabricated on a single chip are quite well matched. An example is shown in Prob. 15.1-4.

15.2 THE CMOS ASTABLE MULTIVIBRATOR

The circuit diagram of a CMOS NOR-gate astable multi is shown in Fig. 15.2-1a. Again, in order that we may describe the essential features of the operation without going afielid in a profusion of small details, we assume an input-output characteristic for the gates, as is shown in Fig. 15.1-3. We shall further neglect protective diodes are ideal. Thus, we assume that the input voltage drop is negligible.

With these simplifying assumptions, it is apparent that V and V_{2o} are complementary. When one is at V_{ss} , the other is at zero, and vice versa. Now let us assume that V_{1i} is above V_T . Then V is at zero and V_{2o} is at a fixed voltage V_{ss} . Hence V_{1i} is heading asymptotically toward zero. When V_{1i} reaches V_T , V will change abruptly to V_{ss} and V_{2o} will change abruptly to zero. When V_{1i} reaches abrupt change in V_{2o} will be transmitted through C to V_{1i} . The consequent downward swing at V_{1i} will be limited at ground because of the clamping action of the protective diode at the input to gate G1. Now V_{1i} is below V_T and is heading asymptotically toward V_{ss} , which is the voltage at V . Altogether there will be a periodic switching back and forth in V_{2o} , V , and V_{1i} , as shown by the idealized waveforms in Fig. 15.2-1b, c, and d. The circuit operation does not, of course, depend on V_T being equal to $V_{ss}/2$. If, however, $V_T \neq V_{ss}/2$, the waveform will not be symmetrical; i.e., we shall have $T_1 \neq T_2$. In general we shall have

$$T = T_1 + T_2 = RC \ln \left(\frac{V_{ss}}{V_{ss} - V_T} + \frac{V_{ss}}{V_T} \right) \quad (15.2-1)$$

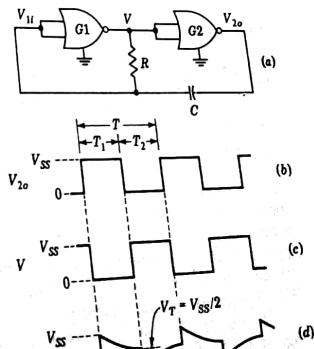


FIGURE 15.2-1

(a) A CMOS astable multi. (b) to (d) Idealized waveforms of the circuit.

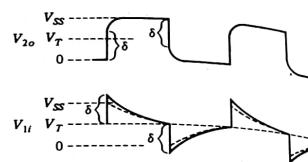


FIGURE 15.2-2
Waveforms of Fig. 15.2-1b and d taking into account the finite output impedance of gate G2 and the use of real diodes.

and for $T_1 = T_2$

$$T = 1.4RC \quad (15.2-2)$$

We should note that the waveforms in Fig. 15.2-1 are actually inconsistent with one another. For we observe that on one side of the capacitor the voltage changes abruptly by amount V_{ss} while on the other side the abrupt changes are of magnitude $V_{ss}/2$. In a physical circuit the voltage $V_{ss}/2$, which does not appear in V_{1u} , actually develops across the output impedance of gate G2 and across the input diode of gate G1. That is, at each transition there is a brief interval during which V_{2o} does not attain V_{ss} or 0 and during that same interval V_{1u} extends above V_{ss} or below ground. During this brief interval the capacitor charges or discharges and thereby changes its voltage by amount $V_{ss}/2$. More realistic waveforms are illustrated in Fig. 15.2-2.

In Prob. 15.2-2 there is shown an alternative form of the astable CMOS multi which has the merit that its timing is less sensitive to variations of V_t and variations of supply voltage.

15.3 MONOSTABLE MULTIVIBRATORS USING ECL GATES

We turn our attention now to monostable circuits using integrated-circuit ECL NOR gates. A flip-flop using ECL gates is shown in Fig. 15.3-1. The flip-flop supplies the required gating functions, and the external r and C supply the required timing.

It will be recalled that ECL gates use transistors of high current gain and further that these transistors are always either in the active region or are cut off, i.e., never in saturation. Each input S (set) and R (reset) is applied directly to the base of a transistor in gate G2 and G1, respectively. The current into a base is zero when the transistor is cut off. And even when the transistor is on, the input current may well be small enough for this input current not to have serious influence on the operation of the circuit. It is these features that make the ECL flip-flop particularly suitable for use in the configuration of Fig. 15.3-1.

In Fig. 15.3-1, when the input at S is at voltage $V(0)$ corresponding to logic 0, the flip-flop will find itself in a permanently stable state with the flip-

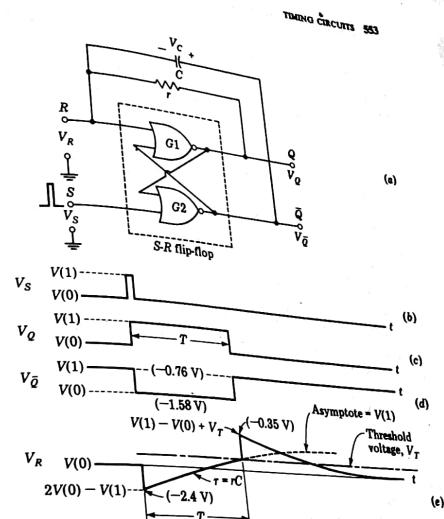


FIGURE 15.3-1
(a) A monostable multi using ECL NOR gates. (b to e) Waveforms of the circuit.

flop in the reset condition, that is, $Q = 0, \bar{Q} = 1$. The student may verify that such is the case by assuming that it is so and then showing that all logic levels at the inputs and outputs of the gates are consistent with one another. Assuming $Q = 1, \bar{Q} = 0$ leads to inconsistencies. In this stable reset condition $V_Q = V_R = V(0)$. The R input of gate G1 looks into the base of a transistor which is cut off. No current flows into this base, and the current through r is also zero. The voltage across the capacitor C is $V_C = V(1) - V(0)$.

Now let the S input rise briefly to the logic level corresponding to voltage $V(1)$, as shown in Fig. 15.3-1b. Then, as appears in Fig. 15.3-1c and d, the flip-flop will make a transition to the set state. The voltage $V_{\bar{Q}}$ will drop abruptly by amount $V(1) - V(0)$. This abrupt change will be transmitted through capacitor C to node R , so that R will now find itself at $V_R = V(0) - [V(1) - V(0)] = 2V(0) - V(1)$, as shown. Note that since the input R has been driven negative, the corresponding transistor is farther in cutoff so that again no current flows into the input R .

The voltage V_R cannot remain at $V_R = 2V(0) - V(1)$ because now we find V_Q at $V_Q = V(1)$. There is a voltage $V_r = V(1) - [2V(0) - V(1)] = 2[V(1) - V(0)]$ across r , and current flows through r , charging capacitor C . Accordingly, the voltage V_R increases exponentially with time constant

$$\tau = rC \quad (15.3-1)$$

heading asymptotically toward $V(1)$. [In writing Eq. (15.3-1) we have ignored the output impedance of gate G_2 because in an ECL gate it is ordinarily very small in comparison with r .] As V_R rises, a point will be reached where $V_R = V_T$, V_T being the transition voltage of the flip-flop. We assume again, for simplicity, that there is a sharply defined voltage $V_R = V_T$ at which the flip-flop makes a transition between states, the voltage levels of the flip-flop being otherwise independent of V_R . At this point, a rapid reversal in the state of the flip-flop will take place back to the reset condition. The voltage $V_{\bar{Q}}$ will jump by amount $V(1) - V(0)$ and there will be an equal jump at point R . We shall then have, as shown, $V_R = V(1) - V(0) + V_T$. It may be verified that the time T of the quasi-stable state is

$$T = \tau \ln 2 \frac{V(1) - V(0)}{V(1) - V_T} \quad (15.3-2)$$

If the R input of gate G_1 continued to draw no input current, V_R would decay from its peak value, $V(1) - V(0) + V_T$, toward its asymptotic limit $V(0)$ with the same time constant τ as prevailed during the quasi-stable state. In order to ensure that successive cycles would duplicate one another it would be necessary to allow a recovery time of several (3 or 4) time constants. As we may anticipate, T itself is of the order of τ . Hence we should have to allow a recovery time which is several times longer than the timing interval itself. This feature, which may well be quite inconvenient, can be corrected in a manner to be noted below. Actually, from the point where V_R reaches its peak value until it falls to V_T the gate input does draw some current, which is supplied through the capacitor C , and thereby the capacitor recharges more rapidly. However, as noted, the gate input current is small and makes no substantial improvement in the recovery time.

As noted in Chap. 7, the logic levels of an ECL gate operating with a supply voltage -5.2 V are $V(0) \approx -1.58$ V and $V(1) \approx -0.76$ V. The threshold voltage V_T is midway between these levels at $V_T \approx -1.17$ V. The other voltages indicated in Fig. 15.3-1 are

$$2V(0) - V(1) = -2.4 \text{ V} \quad \text{and} \quad V(1) - V(0) + V_T = -0.35 \text{ V}$$

Using Eq. (15.3-2), we find the quasi-stable-state duration T to be

$$T = rC \ln 2 \frac{V(1) - V(0)}{V(1) - [V(1) + V(0)]/2} = rC \ln 4 = 1.4rC \quad (15.3-3)$$

It is of interest to note that experimentally, using the Motorola MC 302 ECL flip-flop, T is found to be given by

$$T = 20 \text{ ns} + 1.4r(C + 5 \text{ pF}) \quad (15.3-4)$$

In Eq. (15.3-4), the 20 ns is due to propagation delay in the flip-flop, which we have not taken into account, and the 5 pF added to C is necessary because of the inevitable stray capacitance.

Finally, we may note that the circuit would perform equivalently if instead of as shown in Fig. 15.3-1, C were connected from Q to S and r connected from \bar{Q} to S . In this case the triggering pulse would be applied instead to R . Other types of monostable circuits which employ clocked JK ECL flip-flops are also available commercially.

Triggering While we have not so indicated in the discussion above or in the waveforms of Fig. 15.3-1, the amplitude and duration of the triggering waveform V_S applied at S will have an effect on the duration of the quasi-stable state. Observe that in Fig. 15.3-1a the NOR outputs of the gate are being used. Referring to Fig. 7.4-1b, we note that when the ECL-gate output voltage enters the region corresponding to logic level 1, the NOR output depends on how high the input voltage goes. Thus, more realistically, the circuit waveforms appear as in Fig. 15.3-2. The reflection of V_S in the waveform of V_R will clearly have an effect on the circuit timing. It is therefore advantageous to modify the circuit so that it is unresponsive to the triggering waveform after it has responded to its leading edge. Such a modification is shown in Fig. 15.3-3.

In this figure we have inserted an additional NOR gate G_3 and we have inverted the polarity of the trigger. In the reset state, $Q = 0$ and the trigger level is high, so that the set input is at logic 0. When the trigger drops to 0, the S input goes to 1 and the flip-flop is set. As soon as Q becomes 1, that is,

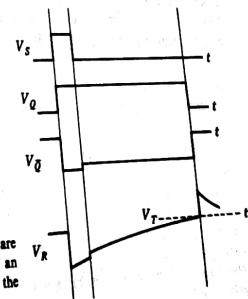


FIGURE 15.3-2

How the waveforms of Fig. 15.3-1 are modified because the NOR output of an ECL circuit varies with input when the input is in the range of logic level 1.

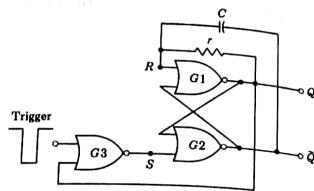


FIGURE 15.3-3
A modification of the circuit of Fig. 15.3-1 to ensure that the triggering waveform does not affect the timing of the multi.

after a time delay equal to the sum of the propagation delays in G_1 and G_2 , gate G_3 is disabled and its output is set to 0 independently of the trigger waveform. Note that the pulse width of the trigger must be at least equal to the sum of the propagation delays in G_1 and G_2 . Observe that the S input voltage is now also independent of the trigger amplitude.

An alternate configuration of an ECL one-shot is shown in Fig. 15.3-4. Here we have replaced the two NOR gates G_1 and G_2 of Fig. 15.3-3 by an SR flip-flop, such as the Motorola MC 1014, and replaced gate G_3 by an edge-triggered JK flip-flop, such as the MC 1013. The virtue of this circuit is that the trigger width can now exceed the pulse width of the monostable multi. It is left for the reader to show that the output-pulse width is independent of the trigger width in the circuit of Fig. 15.3-4 while the output-pulse width must exceed the trigger width when using the circuit shown in Fig. 15.3-3.

Note that triggering of the JK flip-flop occurs on the edge of the input waveform that has a positive slope. If the trigger-pulse rise time is too slow, a Schmitt trigger (Sec. 2.17) must be inserted between the trigger and the JK flip-flop.

Recovery time We have noted that the ECL monostable multi of Fig. 15.3-1 has a long recovery time. Shown in Fig. 15.3-5 are two circuit modifications which reduce recovery time. In Fig. 15.3-5a a diode D has been bridged across

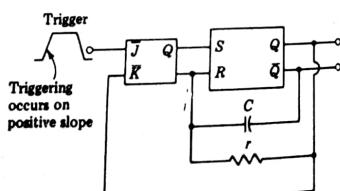


FIGURE 15.3-4
The use of a JK flip-flop rather than a NOR gate (Fig. 15.3-3).

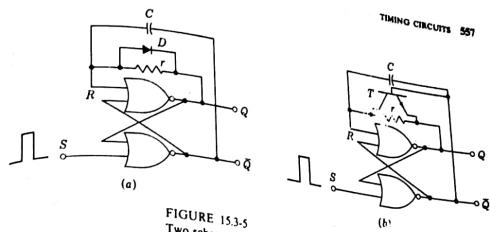


FIGURE 15.3-5
Two schemes to reduce the recovery time in an ECL monostable multi.

r . In the stable state the voltage across r is zero. During the timing interval recovery-time interval the diode is forward-biased, as can be verified from the waveforms of Fig. 15.3-1. When the diode is ON, its incremental resistance r_d will be much smaller than r and hence the capacitor may recharge much more rapidly than in the absence of the diode. However, the diode stops conducting when its voltage falls below about 0.65 V, while the recovery process is not complete until the voltage across r falls to zero. Consequently the diode reduces the recovery time by only about 50 percent.

A much more effective method of reducing recovery time is shown in Fig. 15.3-5b. During the timing interval, as can be verified from the waveform of Fig. 15.3-1, the transistor is cut off. However, during the recovery interval the transistor T is driven vigorously into saturation, thereby effectively short-circuiting the resistor r and discharging C .

15.4 MULTIVIBRATORS FOR SHORT TIMING INTERVALS

When a monostable multi is required to establish a very short timing interval (of the order of tens of nanoseconds), it is more practical to establish the timing interval with a delay line than with a resistor-capacitor circuit. Such delay-line multis generally use ECL logic gates whose propagation delays are comparably short.

A monostable circuit using ECL gates with their NOR and OR outputs is shown in Fig. 15.4-1. We shall assume a propagation-delay time t_{pd} for the gates themselves and assume as well that this delay is the same for the OR and the NOR outputs of gates G_A and G_B . The delay line itself has a delay time t_D .

The operation of the circuit is described by the logic waveforms in Fig. 15.4-1b. For the present we consider only the waveforms drawn in solid lines. We start with V_i at logic level 0 and the circuit in a stable steady state. Then

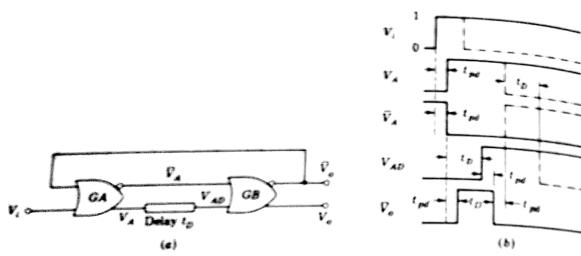


FIGURE 15.4-1
(a) A monostable multi using ECL gates and using a delay line to establish the timing. (b) Logic waveforms of the circuit.

it can be verified that the second input to gate \$GA\$ (which is also the \$GB\$ output \$V_o\$) is also at logic 0. (Assuming that \$V_o\$ is at logic 1 leads to inconsistencies.) Now let \$V_i\$ jump to logic 1, as shown. Then, after a delay \$t_{pd}\$, \$V_A\$, which was at 0, will jump to 1 and \$V_o\$ will change to 0. The logic level of \$V_{AD}\$ is the same as \$V_o\$ except for the time delay \$t_D\$. Finally, \$V_o\$, which is \$V_o = 1\$ whenever both \$V_A\$ and \$V_{AD}\$ are 0, is as shown. In drawing \$V_o\$ we have again taken account of the propagation delay \$t_{pd}\$, this time through gate \$GB\$. We now observe that we have established that \$V_o\$ is \$V_o = 1\$ for exactly the duration \$t_D\$. The interval during which \$V_o = 1\$ begins after a delay \$2t_{pd}\$, but the duration of the interval is \$t_D\$ entirely independently of \$t_{pd}\$.

In Fig. 15.4-1a with \$V_i\$ a step, the feedback connection from the output of gate \$GB\$ to the input of gate \$GA\$ serves no purpose. For as long as \$V_i = 1\$ the output of \$GA\$ is \$V_A = 0\$ no matter what the logic level at the other input. The feedback connection, however, does permit us to replace the step by a pulse, as indicated by the dashed waveforms. For, in the presence of the feedback, \$V_i\$ needs to be held at \$V_i = 1\$ only long enough for \$V_o\$ to rise to \$V_o = 1\$, that is, for a time equal to or longer than the sum of the delays of the two gates, that is, \$2t_{pd}\$. With a pulse input all the waveforms return eventually to their initial levels. As can be verified from the waveforms, if \$V_i\$ changes to \$V_i = 1\$ at \$t = 0\$, the circuit will have completely returned to its initial state at a time \$t = 2t_D + 3t_{pd}\$. The duty cycle of the multi is defined as the maximum fraction of the interval between successive triggering inputs which can be taken up by the quasi-stable state of the multi. In the present case the duty cycle is seen to be \$t_D/(2t_D + 3t_{pd})\$, which is at most 50 percent if \$t_D \gg t_{pd}\$.

An alternative arrangement which allows an improved duty cycle is shown in Fig. 15.4-2. The circuit is recognized as a flip-flop (two crossed coupled inverters) in which the input \$V_i\$, directly or delayed, provides the set and reset

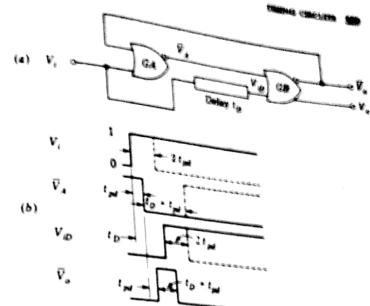


FIGURE 15.4-2
An alternative to the circuit of Fig. 15.4-1 which allows an improved duty cycle.
(a) Circuit diagram. (b) Logic waveforms.

inputs. Hence, initially, with \$V_i = 0\$ two states are possible. We shall assume that, to start, the circuit is in the state in which \$V_o = 0\$, \$V_A = 1\$. As before, we assume equal gate delays \$t_{pd}\$.

The solid waveforms in Fig. 15.4-2b correspond to the case in which \$V_i\$ is a step from logic 0 to logic 1. We then observe that the timed waveform \$V_o\$ has a duration \$t_D - t_{pd}\$ beginning at a time \$2t_{pd}\$ after the jump in \$V_i\$. As in the previous case, so here, because of the feedback, the step in \$V_i\$ can be replaced by a pulse of duration at least equal to \$2t_{pd}\$. The waveforms for a pulse input are shown dashed. As can be seen in the waveforms, the timed output \$V_o\$ is the same for step or pulse input. With a pulse input of minimum duration \$2t_{pd}\$ the waveform returns to its initial state at a time \$t_{pd} + 2t_{pd}\$. The duty cycle is therefore \$(t_D - t_{pd})/(t_D + 2t_{pd})\$, which may approach 100 percent for \$t_D \gg t_{pd}\$.

The ECL astable multivibrator An ECL astable multivibrator can be constructed using the circuit shown in Fig. 15.2-1a by replacing the CMOS gates G1 and G2 by ECL NOR gates.

15.5 AN INTEGRATED-CIRCUIT TTL MONOSTABLE MULTIVIBRATOR

The 54121 TTL integrated-circuit monostable multi is shown in block-diagram form in Fig. 15.5-1a. The circuit is seen to consist of two parts, a flip-flop (gates G1 to G4) and a trigger pulse-shaping circuit (gates G5 and G6). In the actual circuit gates G1, G3, G4, and G5 are TTL gates, each with a logic 1 level of 3.5 V

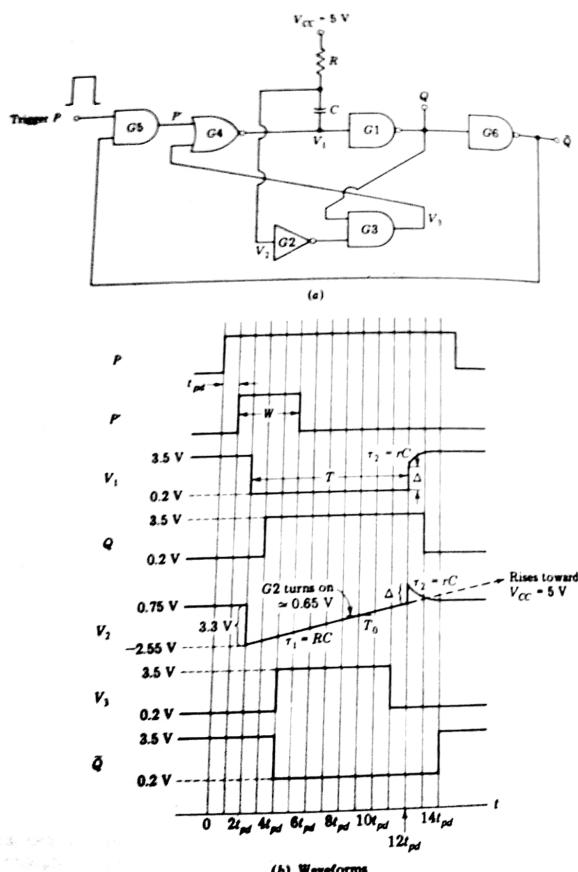


FIGURE 15.5-1
(a) Block diagram of monostable multivibrator. (b) Waveforms.

and a logic 0 level of 0.2 V, while gate G2 is a grounded-emitter amplifier which saturates when its input voltage $V_2 = 0.75$ V and has a cut-in voltage of 0.65 V. In addition, in the actual circuit, the trigger input is first shaped by a Schmitt trigger (not shown) before application to gate G5.

In steady state $P = 0$, and therefore $P' = 0$. Furthermore, in steady state, there is no current flowing through capacitor C. Thus, any current in R flows into the base of the grounded-emitter inverter, gate G2. The resistor R is small enough for G2 to be saturated, and $V_2 = 0.75$ V. Since G2 is saturated, V_3 is at logic 0, and with P also at 0, the output of G4 is at logic 1. We then see that $Q = 0$ and $\bar{Q} = 1$. Gate G5 is thereby enabled, so that P' will rise to 1 when P

Now assume that a positive step is applied at P at time $t = 0$ (see Fig. 15.5-1b). Since G5 is enabled, P' goes high (the delay t_{pd} shown in Fig. 15.5-1b is the propagation delay, which we shall assume to be the same for each gate in the circuit). The output V_1 of gate G4 drops abruptly to 0.2 V. Since V_1 and V_3 are coupled by capacitor C, V_2 decreases by the same amount as V_1 , an amount which is approximately $3.5 - 0.2 = 3.3$ V. Capacitor C now begins to charge, bringing V_2 from $0.75 - 3.3 = -2.55$ V asymptotically toward $V_{cc} = 5$ V. When V_2 rises to approximately 0.65 V, G2 turns on and V_3 goes low. In Fig. 15.5-1b we have shown V_2 reaching 0.65 V at a time T_0 . To simplify the figure we have caused T_0 to be a multiple of the propagation-delay time and have arbitrarily set $T_0 = 9t_{pd}$. The delay between V_2 reaching 0.65 V at time T_0 and V_3 falling is seen to be equal to the delay due to gates G2 and G3. This delay, $2t_{pd}$, is independent of the value of T_0 .

The timing interval of the multi is completed at $t = T_0$. At this time, except for the propagation delays, the circuit returns to its permanently stable state. Suppose now that we want the circuit to respond again to a succeeding trigger and that we want the timing interval in response to this second trigger to be the same as to the first triggering signal. If such is the case, we must allow an interval between triggers for the charge on C to return to its initial value. The recharging of C gives rise to an overshoot in the waveform of V_2 , and correspondingly, the waveform of V_1 does not attain its logic 1 level of 3.5 V until this overshoot is completed. Since the voltages V_1 and V_2 appear at the two sides of a capacitor, the voltage jumps Δ in the two must be equal, as indicated in Fig. 15.5-1. The resistance in series with C as C replaces its charge is, on the one side, the output impedance of G4 and, on the other side, the resistance R in parallel with the input resistance of G2. Let us call this total resistance r. Then the overshoot in V_2 and the approach of V_1 to 3.5 V take place with a time constant $\tau_2 = rC$ (a quantitative calculation of Δ and τ_2 is considered in Prob. 15.5-2).

We also note that when V_1 falls, Q rises and \bar{Q} falls. This causes P' to return to the 0 state. The width W of P' is equal to the sum of the propagation delays of gates G4, G1, G6, and G5 and is independent of P . The width of P must exceed W, however. The virtue of G5 is that as soon as the multi has responded to the input trigger, gate G5 is disabled. The multi is thereby dis-

connected from the triggering source and is free to generate its timing interval without being influenced by the waveform at P . In particular the multi-timing interval T can be less than the interval over which P is held at logic 1.

Output pulse width The circuitry used in the 54121 is designed to minimize variations in output width due to changes in supply voltage and/or temperature. The output-pulse width is nominally

$$T = RC \ln 2 \approx 0.7RC \quad (15.5-1)$$

The output-pulse width T can vary from 40 ns when the external timing components R and C are not used to over 40 s. A duty cycle of 67 to 90 percent is obtainable.

If the supply voltage should vary from 5.25 to 4.75 V (± 5 percent), the pulse width will vary by less than ± 2 percent. In addition, temperature variations from 125 to -55°C result in output-pulse width variations of from +0.2 to -0.8 percent, respectively.

15.6 AN INTEGRATED-CIRCUIT TIMER

In this section we consider an integrated circuit designed to serve in timing applications generally. This integrated-circuit device, type 555, can be used to generate stable time delays like a monostable multi, or it may be used as an oscillator like an astable multi. In the monostable mode, the timing is controlled by one external capacitor and one external resistor. In the astable mode the two operating features of interest, frequency and duty cycle of the output waveform, are controlled by one external capacitor and two external resistors.

The functional diagram of the type 555 timer is shown in Fig. 15.6-1. The names associated with the external terminals are those of the manufacturer. The unit is intended to be used with supply voltages in the range 5 to 15 V. The string of three resistors bridged between the supply voltage V_{cc} and ground provides reference voltages for the two comparators. The reference voltage for comparator 2 is $V_{cc}/3$, and the reference voltage for comparator 1 is $2V_{cc}/3$. As we shall see, these reference voltages have control over the timing. In applications where we may want to vary the timing electronically, we can do so by applying a modulation voltage to the *control-voltage* input terminal. In applications where no such modulation is intended, the manufacturers recommend that the control-voltage terminal be capacitively bypassed to ground (about 0.01 μF).

On a negative-going excursion of the *trigger* input and when the trigger voltage passes through the reference voltage $V_{cc}/3$, the output of comparator 2 sets the flip-flop. On a positive-going excursion of the *threshold* voltage and when the threshold voltage passes through the reference voltage $2V_{cc}/3$, the output of comparator 1 resets the flip-flop. The *reset* input provides a mechanism to reset the flip-flop in a manner which overrides the effect of any instruction to set which the flip-flop may have from comparator 2. This overriding reset will be in effect

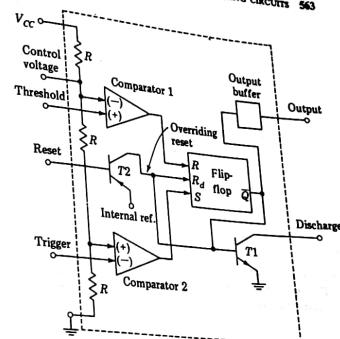


FIGURE 15.6-1
Functional diagram of the type 555 integrated-circuit timer.

whenever the reset input is less than about 0.4 V. When it is not intended that the overriding reset be used, the reset input is ordinarily returned to V_{cc} . The transistor T_2 serves simply as a buffer to isolate the reset input from the flip-flop and the transistor T_1 . The output simply reflects the logic level at the output of the flip-flop. The block marked "output" is a buffer between the flip-flop and the output terminal. This output buffer stage is able to source currents as high as 200 mA and provides logic levels consistent with TTL logic.

An external timing capacitor (not shown in Fig. 15.6-1) may be bridged between the discharge terminal and ground. When the flip-flop is in the reset state, its output \bar{Q} is at logic level 1. Hence, the base of T_1 is high, transistor T_1 is in saturation, and the timing capacitor will be held in a discharged condition. A timing cycle will start when the flip-flop goes to the set state and transistor T_1 is turned OFF. The external capacitor will then be free to charge through an external resistor (not shown). We can terminate this capacitor charging at any arbitrary time by dropping the voltage at the reset terminal below 0.4 V. The direct connection of the override-reset line output of T_2 to the input of T_1 is now seen as a means of turning ON T_1 immediately and thereby circumventing the propagation delay through the flip-flop.

The type 555 timer connected for monostable operation is shown in Fig. 15.6-2. The timing capacitor C is charged from V_{cc} through R . The resistors R_1 and R_2 are rather arbitrary and are of resistance values in the range of many tens of thousands. They are selected to hold the trigger input comfortably above $V_{cc}/3$ in the absence of the negative trigger pulse applied through C .

12

SEMICONDUCTOR MEMORIES

SEMICONDUCTOR MEMORIES 393

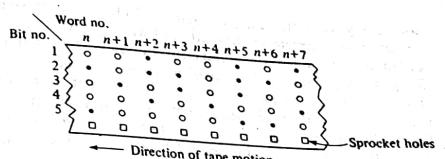


FIGURE 12.1-1
A punched paper tape. An example of a sequentially accessed memory.

to represent logic 1 or logic 0. A logic 1 may be represented by punching a hole in the tape while a logic 0 is represented when no hole is punched in the reserved space. (Of course, if we so choose, the punched hole may represent logic 0, etc.) The bits so recorded across the width of the tape comprise a word, and successive words appear in sequence along the length of the tape. In Fig. 12.1-1 we have made provision for a 5-bit word. The circles represent punchings through the paper and hence are read as logic 1. The dots represent no punching and hence are read as logic 0. Thus, the n th word is 10110, the $(n+1)$ st word 11011, etc. The square holes at the bottom of the tape are sprocket holes used to advance the tape in synchronism with the system of which the tape memory is a part. Words are written onto the tape by mechanically punching holes in the tape as the tape is advanced step by step under the punching head. Words are read in sequence as the tape advances step by step under a reading head. The reading head may consist of five small light sources and an equal number of photosensitive devices.

The essential feature of the memory of Fig. 12.1-1 is that words are written and read in sequence. If, say, the n th word happens to be under the reading head, then the $(n+k)$ th word is not available for reading until after the tape has been advanced k steps. Herein lies a principal limitation of a sequentially accessed memory. If data needed for some purpose are located at random places in a sequential-access memory, a considerable time may be required to assemble the data in one location for processing. On the other hand sequential-access memories have the merit of being relatively inexpensive and are very effective when it is possible to write data into the memory in the order in which they are later to be recalled.

In another type of sequential-access memory, the tape has a coating of magnetic material, and bits are registered on the tape by applying a magnetic field in one direction or the other. As with the punched paper tape, words are written or read as the tape moves under a head. Still other magnetic-coating memories use rotating disks or rotating drums.

In a random-access memory (RAM), words are again stored in locations. Provision is made for singling out a particular location, i.e., addressing the location, for writing data (a sequence of bits) into the addressed location, and

A digital processor generally requires a facility for storing information. The information so stored may consist of the numbers to be used in a computation, intermediate computational results, instructions which will direct a computation, or all three. Where no computation is involved, a storage facility may be called upon simply to store data. For example, a machine designed to address envelopes for mailing will need a storage facility for names and addresses. That part of a digital processor which provides this storage facility is called the *memory*. In this chapter we shall discuss the organization and the electronics of semiconductor memory devices.

12.1 TYPES OF MEMORIES

There are three types of memories: the sequentially accessed memory, the random-access memory (RAM), and the read-only memory (ROM).

A simple and straightforward example of a sequentially accessed memory is the punched paper tape shown in Fig. 12.1-1. The tape is a strip of paper of extended length across whose width at regular intervals are reserved spaces used

for calling forth the data, i.e., reading from the addressed location. The time required to complete the operation of writing a word into a memory location is called the write-access time (and correspondingly for reading the read-access time). Suppose now that we have read a word from, or written a word into, some first location in a memory. Suppose further that we now turn our attention to some second location, chosen arbitrarily, i.e., at random, and access that location for reading or writing. In a random-access memory the access time to this second location is the same for all locations. This situation is different from that prevailing in a sequential memory system, where the access time to a second location depends on its location with respect to the first accessed location.

The read-only memory (ROM) is a type of random-access memory. The ROM differs from the RAM in that in a ROM no provision is made to write words into the memory while the system is operating, i.e., in real time. The content of the memory is usually established by the manufacturer or user and thereafter generally cannot be altered by the user.

12.2 SHIFT-REGISTER SEQUENTIAL MEMORIES

The sequential memories of the previous section, involving mechanically moving parts as they do, are limited in speed. A shift-register sequential memory is shown in Fig. 12.2-1. This memory uses N shift registers, each of K stages, and will provide storage for K words, each of N bits. Each register holds one of the N bits of each of the K words. With each clock cycle (the clock input

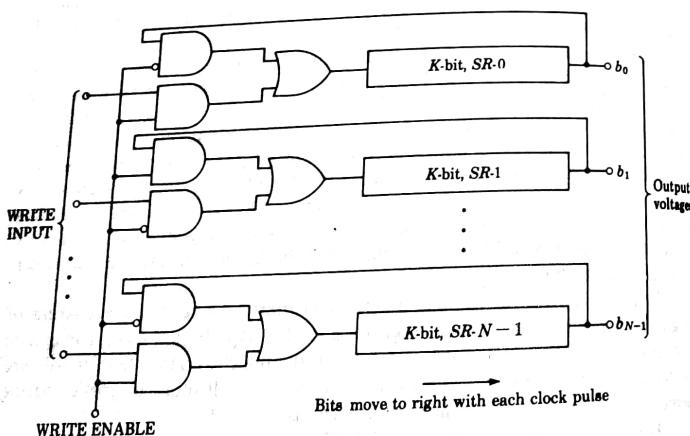


FIGURE 12.2-1
A sequential-access memory using shift registers.

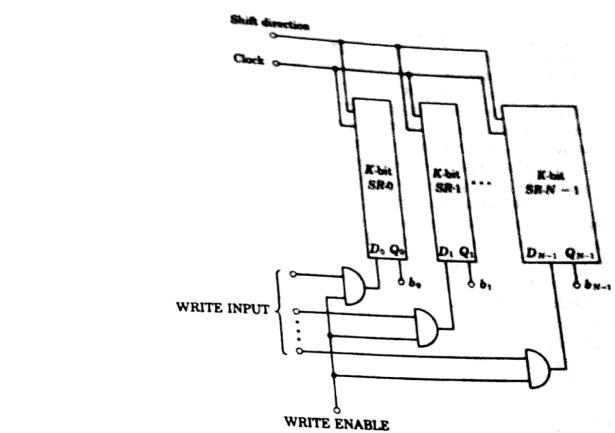


FIGURE 12.2-2
A FILO push-down-stack sequential shift-register memory.

is not explicitly indicated in Fig. 12.2-1) the bits in each register will advance one step, and the stored words will appear sequentially at the outputs b_0, b_1, \dots, b_{N-1} . In this sequential memory the word being read, which is the ensemble of bits registered in the last register position, may be transferred back to the leftmost register position in the manner of a recirculating shift register. Thus, after the entire sequence of words has appeared at the memory output, the same sequence may be repeated endlessly as long as the clock continues to drive the registers. In this sense the present register corresponds to a tape memory in which the beginning and end of the tape have been joined so that the tape forms a closed loop.

If the logic level on the write-enable line is set at logic 1, the recirculation path around the registers will be broken. In this case, the bits applied at the write-input lines, i.e., applied in synchronism with the clock, will successively fill register positions in the shift registers. Thus, with the write control at logic 1 we can erase and replace the contents of the memory.

Starting with the memory initially cleared, the first word written into the memory will be the first word that will appear at the memory output when the content of the memory is recalled. Hence, the present memory is described as a first-in first-out (FIFO) system.

A different mode of organization of a shift-register sequential memory is shown in Fig. 12.2-2. Here again N shift registers each of K bits are used to

provide a memory for N -bit words, the memory having a capacity of K such words. The present memory differs from the memory of Fig. 12.2-1 in three important respects. The circulation feature of the previous circuit is not employed here, and the data input and data output are applied to and taken from the same register stage. Finally the shift registers have the feature that they can be shifted in either direction (see Sec. 10.5). If we think of a typical register stage in the shift register as a type- D flip-flop, then the data input is applied to the D (data) input of the bottom flip-flop in each shift register and the word is recalled by reading the bits of the Q outputs of these same flip-flops.

Let the write enable be at logic 1 and the shift-direction line set to shift bits upward, and let a sequence of words be presented at the write-input lines in synchronism with the clock. Then a total of K words can be stored in the memory. Each word enters at the bottom and moves up one stage to make room for the next word. Figuratively, the words are stacked one on top of another, the first input word being at the top of the stack. To read the remembered words in sequence we set the write enable at logic 0 and reverse the direction of shifting to the downward direction. With each clock cycle a remembered word appears at the memory output in the order opposite to that in which the words were entered. The present memory is hence referred to as a first-in last-out (FIFO) system. Also, we may fancifully picture the operation as one in which a stack of words is pushed down from the top, thereby causing word after word to be squeezed out from the bottom. Such considerations account for the description of the present memory as a *push-down stack*.

12.3 MOS REGISTER STAGES

As noted, MOS technology is of special advantage in large-scale integration (LSI) because an MOS device generally occupies much less real estate on a silicon chip than a comparable bipolar device. Hence, MOS devices are widely used in the large-capacity sequential memories which employ shift registers. We shall consider now the characteristics and operating features of the MOS register stages which are cascaded into shift registers.

We have seen that shift registers are cascades of flip-flops and that these flip-flops must have the operating feature characteristic of the master-slave flip-flop or other equivalent flip-flops, as discussed in Chap. 9. That is to say, at the time in the clock cycle when the output of the flip-flop is responding to input data already present, any change in input data must not affect the output. In considering MOS shift-register stages, we may be hard-pressed to know whether they are to be called master-slave flip-flops, but we shall see that, within limitations, they accomplish the same function.

Consider the MOS inverter circuit of Fig. 12.3-1. In discussing this and subsequent circuits we shall arbitrarily introduce some symbols which will simplify the circuit diagrams and avoid excessive words. First we use the simplified symbol for an FET which omits the substrate and does not indicate

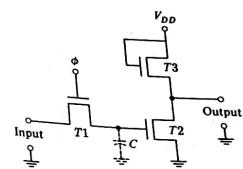


FIGURE 12.3-1
An MOS circuit which can store a bit temporarily.

whether the device is a PMOS or an NMOS unit; however, we shall assume throughout this chapter that NMOS is being employed. In this case the supply voltage V_{DD} is positive, and an FET will be conducting when its gate voltage is positive and will be cut off when the gate voltage is zero (at ground). We assume, further, positive logic, so that an FET is conducting when its gate is at logic 1 and not conducting when its gate is at logic 0, which we shall take to be nominally ground.

Returning now to the circuit of Fig. 12.3-1, let the voltage ϕ on the gate of T1 go to logic 1 briefly. (We assume here, as in our subsequent discussions, that logic 1 is a high enough voltage to ensure that when a gate is at this level, the FET is ON even if its source is not at ground.) Then the capacitor C will charge to the logic level on the input data line during the interval when $\phi = 1$. (We assume the data line holds fixed during this interval.) The transistor T1 is serving as a switch, called a *transmission gate* (see Sec. 13.5) to connect C to the input data line.

When T1 turns OFF, the inverter formed by T2 and T3 will remember the sampled data because of the charge stored on the capacitor. The capacitor is shown dashed to indicate that it represents stray, incidental capacitance present at the gate of a transistor and not a capacitor deliberately introduced. Typically C may be in the range of 0.5 pF. Since the present circuit stores or remembers a bit of data, it serves the same function as a flip-flop. However, in comparison with the flip-flop the present circuit has the limitation that its memory is short-lived. The charge on C will eventually leak off. This loss of charge is due in some small extent to leakage through the insulation which supports the gate on the transistor T2. To a much larger extent it is due to the leakage through the reverse-biased junction formed between the substrate and the drain of T1. Typically, it turns out that the capacitor is able to hold adequate charge for an interval which is of the order of magnitude of about 1 ms. If bipolar devices were used in the present circuit, the storage time would be many orders of magnitude smaller because of the relatively large base current required to drive such a bipolar transistor. With bipolar transistors the storage time is so small that the present circuit is generally entirely ineffective.

A shift-register stage can be constructed, as shown in Fig. 12.3-2, from

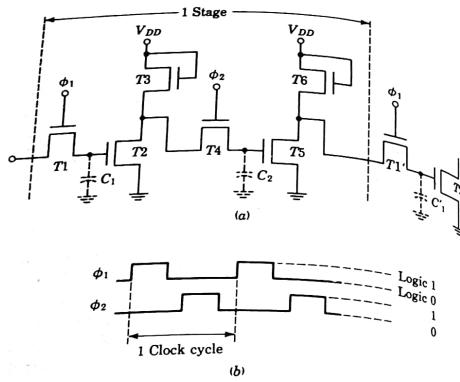


FIGURE 12.3-2
An MOS dynamic shift-register stage.

two inverter stages as in Fig. 12.3-1. Two gating voltages ϕ_1 and ϕ_2 for transmission gates $T1$ and $T4$ are required. These gating waveforms, as shown in Fig. 12.3-2b, are phased such that $T1$ and $T4$ will not be ON simultaneously. When $\phi_1 = 1$, the input data bit to the stage is transferred to C_1 . At the same time the complement of the bit stored on C_2 is transferred to C_1 , the input storage capacitor of the next register stage. While $\phi_1 = 1$, $\phi_2 = 0$; hence, while the new input bit is being stored on C_1 and possibly changing the output of the input inverter ($T2$ and $T3$), this new input bit will not affect the output stage or the bit being transferred to the next stage. When ϕ_2 becomes $\phi_2 = 1$, the complement of the bit on C_1 will be transferred to C_2 . The similarity between the operation of the present circuit and the operation of a master-slave flip-flop is apparent. The first inverter stage with its storage capacitor serves as the master flip-flop, and the second inverter stage with its storage capacitor serves as the slave flip-flop. The fact that two phases of gating voltage are required is to be compared with the fact that in the master-slave flip-flop the clock waveforms for master and slave are complementary waveforms.

It must be remembered, however, that the circuit shown in Fig. 12.3-2 is not really a flip-flop. For, if the time between ϕ_1 going to the 0 state and then to the 1 state is too great, the charge stored on capacitor C_1 will have leaked off. To avoid this possibility the shift register is operated "dynamically," and the maximum period of ϕ_1 (and ϕ_2) is typically less than 1 ms. Thus, this type of shift register is referred to as a *dynamic shift register*.

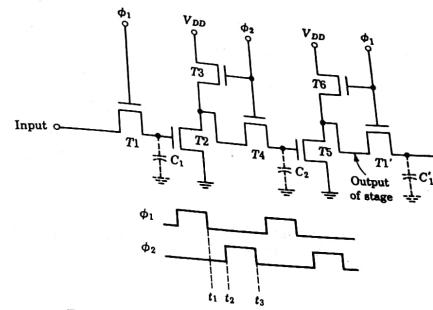


FIGURE 12.3-3
An MOS dynamic shift-register stage with load-transistor clocking.

In the circuit of Fig. 12.3-2, if the bit stored on C_1 is 1, then during an entire bit interval $T2$ and $T3$ will be conducting, dissipating power, and drawing current from the supply V_{DD} . If the bit is a 0, then $T3$ and $T4$ will not conduct but there will then be an interval equal to the duration of a clock cycle during which C_2 will be storing a 1 and during which $T5$ and $T6$ will conduct and dissipate power. Power dissipation can be significantly reduced through the circuit modification shown in Fig. 12.3-3. Here the gates of the inverter are not held at V_{DD} but are clocked so that $T3$ conducts only when $T4$ conducts and $T6$ conducts only when $T1'$ conducts. If the bit stored on C_1 or C_2 is a 0, then the corresponding inverter will not conduct. If, however, a 1 is present on C_1 or C_2 , then the corresponding inverter will conduct, but only during the time interval when either ϕ_1 or ϕ_2 is at logic 1. The waveform ϕ_2 (or ϕ_1), however, need be held at logic 1 only long enough to allow C_2 (or C_1) to charge from V_{DD} through $T3$ and $T4$ (or through two other corresponding transistors). It turns out that this charging time is generally appreciably smaller than a full clock cycle (especially at low clock rates), and hence, the *clocked-load* arrangement of Fig. 12.3-3 can effect a good reduction in power dissipation. With the clocked load, however, the capacitive load on the clock driver is increased, and in the fabrication of the integrated-circuit chip there is the increased topological complexity occasioned by the need to distribute the external clock to all the load gates. As a result of the large current requirement, it often turns out to be necessary that the clock driver be a TTL circuit.

A variation of the clocked-load register stage is shown in Fig. 12.3-4. Here the clock for the transmission gates ($T1$, $T4$, etc.) is applied to the load gate of the succeeding inverter rather than, as in Fig. 12.3-3, to the load gate of the

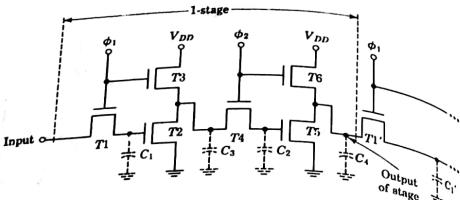


FIGURE 12.3-4
An alternative MOS dynamic shift-register stage with load-transistor clocking.

preceding inverter. In the present case the circuit operation requires the presence of an additional capacitor C_3 at the output of the inverter. When $\phi_1 = 1$, the input data bit is transferred to C_1 and its complement to C_3 . When $\phi_2 = 1$, the bit on C_3 is transferred to C_2 and its complement to C_4 . The input data bit is now stored on C_4 which is the input of the next stage.

To see the need for C_3 consider again the transfer of the input data bit through the first inverter and through transmission gate $T4$ to C_2 . Assume that the input bit is a 0 and that C_3 is not present. Then when ϕ_1 becomes $\phi_1 = 1$, the complement of the input data bit will appear at the output of the inverter. But when ϕ_1 returns to $\phi_1 = 0$, both $T3$ and $T2$ become nonconducting and the inverter output is isolated from its input. If, however, C_3 is present, then C_3 can store the input data complement. Then when ϕ_2 goes to $\phi_2 = 1$, the logic level 1 can be transferred to C_2 through $T4$. Since C_3 must charge C_2 and still maintain a voltage which is comfortably in the range of logic 1, then, as a matter of fact it is required that $C_3 \gg C_2$. As can be verified, for the case where the input data bit is a 1, the presence of C_3 is not required.

12.4 TWO-PHASE RATIOLESS SHIFT REGISTER

The shift-register stages of the preceding section involve, in every case, straightforward inverter stages. As was discussed in Sec. 8.3, such stages are composed of a driver FET and a load FET. It is required that there be a large ratio between the channel resistance of the driver and the channel resistance of the load, that is, $(W/L)_D \gg (W/L)_L$, to ensure a narrow transition region. As was also noted, such ratio-type inverter stages are obtained by making the load-FET channel long and narrow in comparison with the driver-FET channel. Correspondingly, the load FET occupies a large area of the integrated-circuit chip in comparison with the area of the driver. Further, because of the high

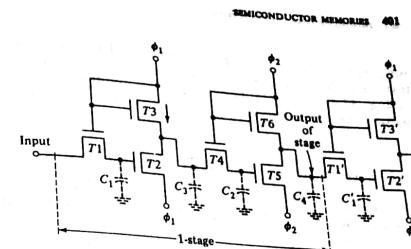


FIGURE 12.4-1
A ratioless dynamic shift-register stage.

resistance of the load channel, the speed of operation of the inverter is limited. For in any complete cycle of operation there will be an interval when some capacitance will have to charge through this high resistance. Both these disadvantages, in the matter of area and operating speed, can be relieved by devising a scheme which allows the inverter to be replaced by two low-resistance FETs of the same geometry.

Such a ratioless shift register stage is shown in Fig. 12.4-1. Observe that this circuit is similar to that shown in Fig. 12.3-4 except that both supply voltage and ground connections have been replaced by connections to the clock phases. In the ensuing discussion it will be convenient to think that $\phi = 0$ means that ϕ is at 0 V and that when $\phi = 1$, ϕ is at some positive voltage V_{DD} (say 10 V). Similarly, let the data to be transferred through the register stage have logic levels 0 and 1, which are 0 and V_{DD} , respectively.

Now let the input data bit be at V_{DD} . Then when ϕ_1 goes to V_{DD} , C_1 will charge to V_{DD} . Since with ϕ_1 at V_{DD} , $T3$ will also be ON and, $T3$ will charge capacitor C_3 to V_{DD} . Note that during the operation of charging C_3 , transistor $T2$ is OFF, since at no time does the gate voltage of $T2$ exceed the source voltage of $T2$ by the threshold voltage V_T . Thus the combination of $T2$ and $T3$ does not operate in the manner of an inverter.

When ϕ_1 returns to 0 V, $T3$ goes OFF and $T2$ goes ON because the charge stored on C_1 causes $V_{GS}(T2)$ to exceed the threshold voltage. As a result C_3 discharges through $T2$, and the capacitor C_3 is left at 0 V. The overall result is that after the clock pulse, ϕ_1 has returned to 0 V, the bit stored on C_3 will be the complement of the input bit. Hence, the net effect is the same as if the combination $T2$, $T3$ were indeed an inverter. It can similarly be verified that if the input data bit were a 0, an inversion would again take place. In a similar

way, when ϕ_2 goes through its cycle from 0 to 1 and back to 0, it will have transferred to C_4 the complement of the bit on C_3 . The ratioless register stage of Fig. 12.4-1 places a heavy capacitive burden on the clock signal source because the clock must supply current to charge the capacitors.

12.5 FOUR-PHASE RATIOLESS REGISTER STAGE

In the circuit of Fig. 12.4-1 (as in the circuit of Fig. 12.3-4) we require that $C_3 \gg C_2$ since C_2 must charge from C_3 without causing appreciable voltage change. The large capacitance of C_3 has a twofold disadvantage: (1) it places a limitation on the speed of operation and (2) it requires that, in fabrication, a relatively large area on the integrated-circuit chip be devoted to that capacitor. The scheme shown in Fig. 12.5-1a avoids this difficulty because it does not require that any capacitor take its charge from another capacitor. It has the inconvenience, however, of requiring a four-phase clock waveform. The clock

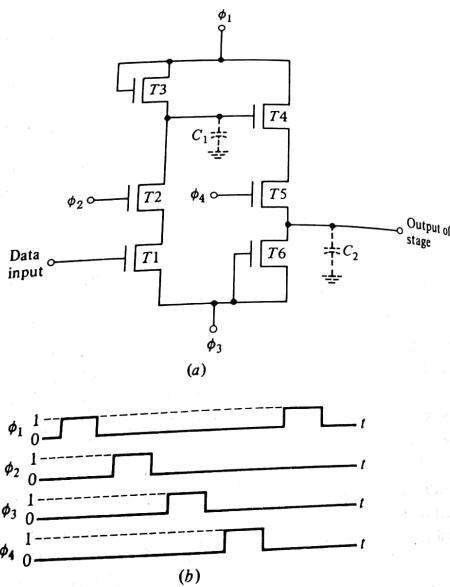


FIGURE 12.5-1
(a) A four-phase ratioless register stage. (b) Clocking waveforms.

waveforms are shown in Fig. 12.5-1b. During the interval when $\phi_1 = 1$, C_1 charges to logic 1. Such is the case, no matter what the input, because $\phi_2 = 0$ and therefore T_2 is OFF. This logic level 1 remains stored on C_1 after ϕ_1 returns to $\phi_1 = 0$. Assume now that the input is at logic 1. Then, when ϕ_2 becomes 1 and turns T_2 ON, C_1 will discharge through T_2 and T_1 , leaving C_1 at logic 0. Thus, we now have stored on C_1 the complement of the input bit. It may similarly be verified that if the input bit is 0, then when ϕ_2 becomes 1, T_1 remains cut off and the bit stored on C_1 will again be the complement of the input bit. In a similar manner, through the operations of the clock phases ϕ_3 and ϕ_4 , the complement of the level stored on C_1 will be transferred to C_2 and hence to the input of the next register stage. This circuit suffers from the disadvantage that ϕ_1 must supply the current to charge C_1 and ϕ_3 must charge C_2 .

12.6 CMOS REGISTER STAGES

Like MOS register stages, CMOS register stages are constructed from inverter stages and transmission gates (taking advantage as well of stray capacitances). The CMOS inverter was discussed in detail in Secs. 1.15 and 8.4 and is shown again in Fig. 12.6-1. For the convenience of being specific, let us assume a supply voltage $V_{SS} = +10$ V. In such a system the input V_i and output V_o will be waveforms which make excursions between 0 V (ground) and 10 V. Let us then turn our attention to the matter of a transmission gate intended for use with voltages in this 0- to 10-V range. A clock waveform is required when a transmission gate is employed. We shall assume that the clock ϕ also makes excursions between 0 V and 10 V.

Consider, first, the MOS transmission-gate arrangement of Fig. 12.6-2a using a single *N*-channel device. The gate is being used to isolate the capacitor from, or to connect the capacitor to, an input V_i which is $V_i = 0$ V or $V_i = 10$ V. Assume that the threshold voltage of the transistor is $V_T = 2$ V. Starting with $V_i = 0$ V and the capacitor voltage $V_C = 0$ V, suppose that V_i becomes $V_i = 10$ V but that $V_G = 0$ V. Then the transistor will remain OFF since

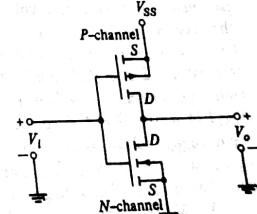


FIGURE 12.6-1
The CMOS inverter.

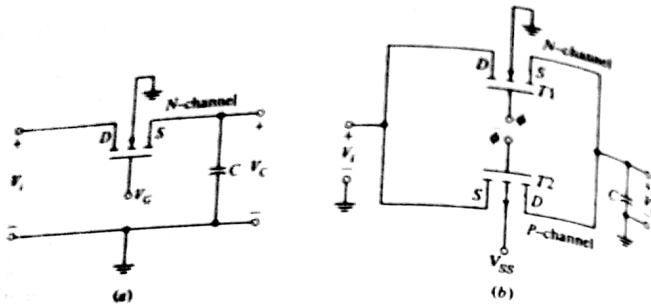


FIGURE 12.6-2
(a) An MOS transmission gate used to charge a capacitor from a voltage source.
(b) A CMOS transmission gate.

$V_{GS} < V_T$ and the capacitor will not charge. Now let V_G become $V_G = 10. Then current will flow from left to right through the transistor. Since the transistor is an *N*-channel MOS, the current direction establishes that the left side of the transistor is the drain and the right side is the source, as indicated in the figure. Since the threshold voltage (which is the gate-to-source voltage V_{GS} at cutoff) is 2 V, the transistor will stop conducting when V_C becomes 8 V since at that point $V_{GS} = V_G - V_C = 10 - 8 = 2. If now, with V_G still at 10 V, the input should become $V_i = 0, the current direction will reverse, the source and drain will reverse, and the capacitor will be able to discharge completely to 0 V. Altogether we have the result that in this system with logic levels separated by 10 V, our single-transistor transmission gate is able to transmit only an 8-V change. Such a limitation applies to all the transmission gates employed in the dynamic MOS register stages discussed in the preceding section. This characteristic, of course, does not preclude the gate from being effective or the register stages from operating, but it does reduce noise margins.$$$

The limitation of the MOS transmission gate is remedied in the CMOS gate, shown in Fig. 12.6-2b. Here the two transistors of a CMOS pair are paralleled, and the gating voltages applied are complementary. Return now to the case where $V_i = 10 and the capacitor is in the process of charging. Current through both transistors flows again from left to right. But since transistor T_2 is a *P*-channel MOS, its source is at the left and its drain to the right, as indicated. When $\phi = 10 and $V_C = 8, we have, as before, $V_{GS}(T_1) = 10 - 8 = 2, so that T_1 is at the point of cutoff. However, for T_2 we have $V_{GS}(T_2) = 10 - \bar{\phi} = 10 - 0 = 10. Since $V_T = 2, transistor T_2 will be ON as long as $\bar{\phi}$ is less than V_i by more than 2 V. Thus, the capacitor can charge to the full voltage $V_C = V_i = 10 since T_2 remains on.$$$$$$$

A CMOS register stage is shown in Fig. 12.6-3. It consists, like the most

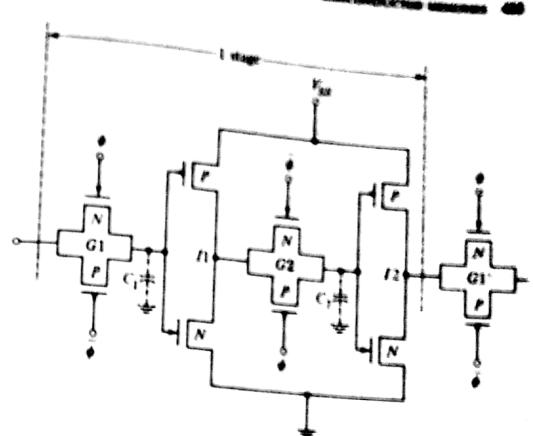


FIGURE 12.6-3
A CMOS register stage consisting of two transmission gates and two inverters.

elementary of the MOS register stages (Fig. 12.3-2), of two transmission gates G_1 and G_2 , two inverters I_1 and I_2 , and the two capacitors C_1 and C_2 . In this figure we have again omitted the substrate connections for simplicity. Observe particularly the reversal of gate phases at alternate gates. The functional operation of this CMOS stage is so similar to the register stage of Fig. 12.3-2 that we need not repeat the description.

We noted that the MOS stage of Fig. 12.3-2 suffered disadvantages due to the fact that the inverters were ratioed and inverter stages would drain a direct current from the V_{SS} supply whenever the inverter gate was at logic 1. Thereafter we considered a series of modifications designed to remedy these difficulties. We may, then, note that the CMOS inverter is "ratioless" and does not draw a steady current from the V_{SS} supply. On the other hand, a CMOS inverter and a CMOS gate use more area on a silicon integrated-circuit chip than a MOS inverter or gate.

12.7 STATIC SHIFT-REGISTER STAGE

The dynamic register stages of the previous section will not operate if the clock interval is longer than the interval over which the stray capacitance is able to store a charge. When a shift register must operate at very low frequencies the dynamic stages must be replaced by static stages.

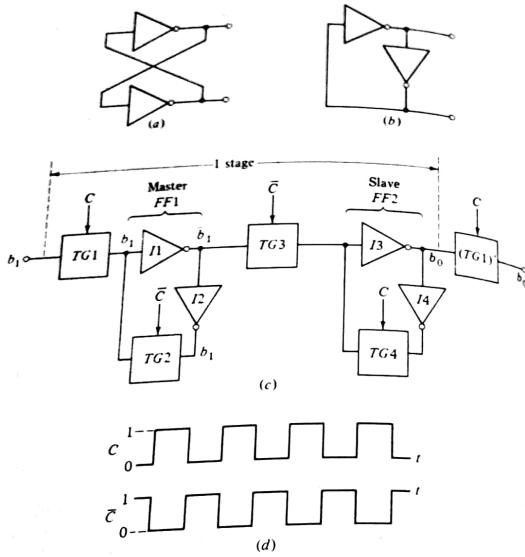


FIGURE 12.7-1
(a) A basic flip-flop with two cross-coupled inverters. (b) The basic flip-flop redrawn. (c) An MOS or CMOS static shift-register stage. (d) The clocking waveforms required in (c).

A conventional flip-flop reduced to its essentials is shown in Fig. 12.7-1a. It is redrawn in Fig. 12.7-1b to make it apparent that the circuit consists of a cascade of two inverters with the output fed back to the input. The circuit of a static shift-register stage is shown in Fig. 12.7-1c. Similar MOS or CMOS circuits were discussed in Sec. 9.21. Note that we have here a master and a slave flip-flop modified by the introduction of transmission gates ($TG1$ and $TG3$) interposed between flip-flops. The circuit diagram of the gates ($TG1$ and $TG3$) is given for the CMOS in Fig. 12.6-1, and the diagram of inverter stages has been given for the CMOS in Fig. 12.6-2b. A MOS transmission gate and inverter is shown in Fig. 12.3-1. Following conventional

practice, the clocking waveform and its complement are here given as C and \bar{C} . These clocking waveforms are shown in Fig. 12.7-1d. The clock inputs of the gates have been labeled to make it clear that $TG1$ and $TG4$ open and close at the same time as $TG2$ and $TG3$. And, of course, when $TG1$ and $TG4$ allow transmission, $TG2$ and $TG3$ do not and vice versa. We shall assume that the clock pulses are applied to the gates in such manner that when $C = 1$ ($\bar{C} = 0$), the gates allow transmission (are short circuits between input and output) and do not allow transmission when $C = 0$. Viewing the gates as switches, we can say that the switch (gate) is CLOSED when $C = 1$ and the switch (gate) is OPEN when $C = 0$. Furthermore while only one clock pulse is shown clocking each transmission gate, it must be understood that in a CMOS system both C and \bar{C} are present.

Let us start at a time when $C = 1$. Then the input data bit b_1 is being presented to $FF1$, but this bit will not be stored in $FF1$ because its feedback path is not closed. The feedback path of $FF2$ is closed, and there is a bit b_0 in storage there. This bit, whose logic level depends on the past history, is being presented through $TG1'$ to the first flip-flop of the next register stage. Note also that $TG3$ is open so that $FF2$ is not influenced by anything to the left of $TG3$. Now let C go to $C = 0$. Then $TG1$ opens. The gate $TG2$ closes, completing the flip-flop circuit and putting the bit b_1 in storage in $FF1$. The bit b_1 at the output of inverter $I2$ has been transmitted through $TG2$ and replaces the previous bit b_0 , which was originally transmitted through $TG1'$. We may wonder whether any timing difficulty might arise because of the finite propagation time through $TG2$. No such problem arises because the inevitable stray capacitance at the input to $I1$ will hold the bit b_1 long enough to await the transmission through $TG2$. Note that here we depend on the stray capacitance to hold a charge only for an interval equal to the gate propagation time and not for the interval of a clock cycle. Hence the clock frequency has no bearing on our present discussion.

At the moment C becomes $C = 0$ and the bit b_1 is being stored in $FF1$, $TG3$ closes, and, after a propagation time (through $TG3$ and $I3$), the bit b_1 appears as a new bit at the input to $TG1'$. But at this time $TG1'$ is OPEN. When C returns to $C = 1$, the feedback path in $FF2$ will close and the bit will remain in storage in $FF2$.

The stage has two important properties relevant to our discussion of registers: (1) if the clock should stop at any point, the bit in the stage will remain in storage, either in $FF1$ or in $FF2$, depending on whether the clock should stop at $C = 0$ or $C = 1$; and (2) as is required generally when register stages are cascaded, the output of one stage changes at that point in the clock cycle when the input gate switch of the next stage is OPEN, thereby disallowing transmission.

12.8 A THREE-PHASE STATIC REGISTER STAGE

In addition to transmission gates, the static shift register of Fig. 12.7-1 has two flip-flops, $FF1$ and $FF2$, each involving two inverters. A static register stage still suitable for service in a shift register can be constructed which uses just a

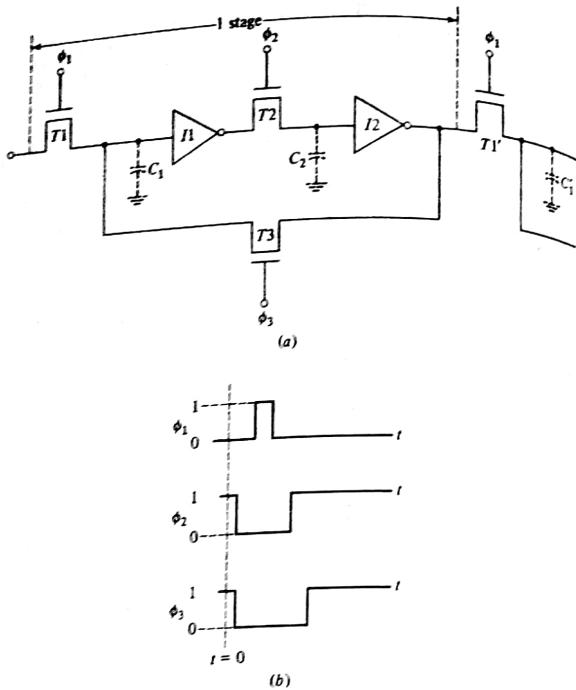


FIGURE 12.8-1
(a) A three-phase static register stage. (b) Clocking waveforms.

single flip-flop, albeit with some added complication in clocking. The circuit is shown in Fig. 12.8-1a. Here, for simplicity, we have shown the transmission gates as MOSFETs. However, CMOS gates can also be employed. The transistors T_1 , T_2 , and T_3 in the present circuit are being used as transmission gates. As before, we assume that the gate allows transmission, i.e., the switch is CLOSED, when the corresponding clock signal is at logic 1 and the switch is OPEN when the clock is at logic 0. A three-phase clock is required.

Consider, then, as shown in Fig. 12.8-1b, that at $t = 0$, $\phi_1 = 0$ while $\phi_2 = \phi_3 = 1$. Then the circuit loop, involving the inverters I_1 and I_2 , is closed and constitutes a flip-flop. Since the two transmission gates T_1 and $T_{1'}$ are open, this flip-flop is isolated from the preceding and succeeding stages. There will be a bit registered in the flip-flop; whether it is a 1 or 0 will be determined by the past history. Suppose that while ϕ_1 remains at $\phi_1 = 0$ we briefly set

$\phi_2 = \phi_3 = 0$. Then the bit in storage in the flip-flop will be preserved on capacitors C_1 and C_2 . If we return $\phi_2 = \phi_3 = 1$ before the charges on C_1 and C_2 have changed appreciably, the flip-flop state will not have changed.

Now let ϕ_2 and ϕ_3 go to logic 0, as in Fig. 12.8-1b. (In the figure it is indicated that ϕ_2 and ϕ_3 go to logic 0 simultaneously. This feature is not essential; nor is it important which clock signal goes to 0 first.) Suppose that shortly after ϕ_2 and ϕ_3 both become 0, ϕ_1 becomes $\phi_1 = 1$. Then because of the voltage held on C_2 the state of the flip-flop, i.e., the logic level held at the output of I_2 , will be transferred to the capacitor C_1 of the next flip-flop. Similarly, the state of the preceding flip-flop will be transferred to C_1 . Now let ϕ_1 return to 0 and bit ϕ_2 be first to go back to logic level 1, thereby closing switch T_2 . Then whatever may have been the previous state of C_2 it will now go to the state complementary to C_1 . Finally, when ϕ_3 goes back to $\phi_3 = 1$, the flip-flop circuit loop is again complete and the flip-flop will hold its new state indefinitely. The order of the reclosing of switches T_2 and T_3 is important. For, as can be verified, if ϕ_3 recloses first, the flip-flop will revert to its original state.

12.9 THE READ-ONLY MEMORY

A read-only memory (ROM) is a memory device intended to store information which is fixed. That is, there is an initial operation during which information is written into the memory and thereafter the memory is *read only* and is not again written into. Generally the information in the memory is placed there by the manufacturer of the device. There are, however, memories which allow the user to establish the store of information in the memory. Such memories are referred to as programmable memories (PROM). There are also ROMs in which the stored information can be changed. However, in such cases the writing operation is accomplished in a time which is many orders of magnitude larger than the time required to read. Such memory devices (described as *erasable*) are read-only memories in the sense that to change the stored information it is necessary to interrupt the digital processing in which the memory is involved.

A most important attribute of the ROM is that the information it stores will not be lost if the electric power that it uses to operate should be interrupted. Such memories are referred to as *nonvolatile*. In contrast, the sequential memories already discussed and the random-access memories to be discussed (beginning in Sec. 12.13) are both *volatile* memories.

The ROM has many applications in a digital system. It can be used to provide the realization of an arbitrary truth table. Whenever a truth table involves enough input and output logical variables for its physical implementation to require a large number of gates, a ROM may well be more economical in size, weight, and cost. ROMs are widely used in code conversion and in connection with alphanumeric displays. The ROM is also used to yield results

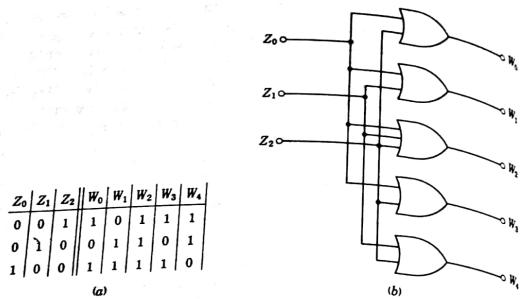


FIGURE 12.9-1
(a) A truth-table specification of an encoder. (b) The encoder realized with OR gates.

that would otherwise be achieved by a computation involving a sequence of arithmetic operations, e.g., multiplication, division, or evaluation of a trigonometric function (as $\sin x$ or $\ln x$), or it may be used as a function generator.

A ROM is an encoder. An encoder is a logical-gate structure which has M inputs Z_0, Z_1, \dots, Z_{M-1} and K outputs W_0, W_1, \dots, W_{K-1} . It is intended that at any time an individual input, say Z_i , is to be singled out, we set $Z_i = 1$ while all other inputs are at logic 0. (Alternatively we may have $Z_i = 0$ while all other outputs are at logic 1.) Corresponding to each input Z_i , which may be set to $Z_i = 1$, the K outputs will take on the logic levels $W_0(i), W_1(i), \dots, W_{K-1}(i)$. That is, the encoder accepts the input Z_i (only) = 1 and identifies this situation through the code word $W_0(i), W_1(i), \dots, W_{K-1}(i)$; or if the encoder is viewed in its application as a memory, the i th storage location in the memory is addressed by setting Z_i (only) = 1 and the ROM responds by presenting at its output the word stored in that location. A logical structure of an encoder is illustrated in Fig. 12.9-1. The truth-table specification is given in Fig. 12.9-1a, and its realization in logic gates is shown in Fig. 12.9-1b. The ROM illustrated stores three words, each of 5 bits.

Rather generally the address of a stored word is itself available in a digital system as a binary coded word. It is therefore necessary to interpose between the binary-coded address word and the ROM a device which takes note of the address and singles out a corresponding individual line. Such a device performs a function which is inverse to the function performed by the encoder and is hence called a decoder. A truth table for a decoder is given in Fig. 12.9-2a. The address word has 2 bits; hence four address words are available. A realization of this decoder with logical gates is shown in Fig. 12.9-2b. Note that in Fig. 12.9-1 only three of the four possible Z_i are being used.

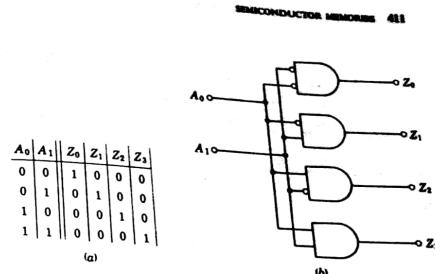


FIGURE 12.9-2
(a) A truth table for a four-word decoder. (b) The decoder realized with AND gates.

As a convenience to the user, manufacturers generally provide decoders as an integral adjunct of the ROM encoder. Such a memory device with decoder included, as indicated in Fig. 12.9-3, are generally referred to as a decoded ROM. In providing the decoder, the manufacturer does himself a service as well. For an integrated-circuit chip there is generally a limitation to the number of pins which can conveniently be provided for external connection. And on an M -word memory without decoder, M pins would have to be provided for addressing purposes; while with the decoder N pins, with $2^N = M$, are adequate.

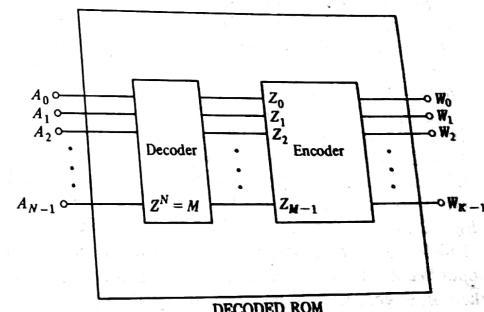


FIGURE 12.9-3
A ROM with decoder included.

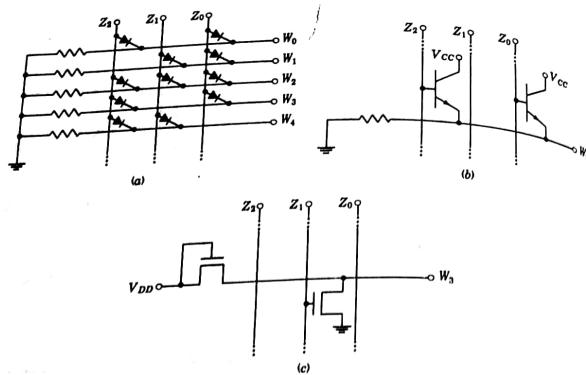


FIGURE 12.10-1
(a) A three-word, 5-bit ROM using diode connections between address and bit lines.
(b) Single-bit line of the ROM in (a) illustrating the use of bipolar transistors in place of diodes.
(c) The use of MOS transistors illustrated.

12.10 IMPLEMENTATION OF ROMs

A possible implementation of the ROM OR-gate encoder shown in Fig. 12.9-1b is shown in Fig. 12.10-1a. It can be verified (Prob. 12.9-2) that the circuit shown does indeed constitute precisely the OR-gate logic structure. The use of diodes, however, has a drawback in that the inputs to the encoder must directly supply the current which must be delivered at the outputs. This difficulty is relieved by using bipolar transistors in place of diodes in the manner shown in Fig. 12.10-1b. Here, for simplicity, we have indicated only the connections to bit output line W_3 . Now the output current may be supplied principally by the V_{CC} source. Since, as we have noted, in bipolar technology, integrated-circuit diodes are generally transistors with collector and base terminals joined, fabrication in the manner of Fig. 12.10-1b is not substantially more complicated than the pattern in Fig. 12.10-1a. It is to be observed further that all the transistors connected to an input line Z_i will have their collector terminals connected in common to V_{CC} and their bases connected in common to the Z_i line. Hence such a vertical array of transistors is not infrequently replaced by a single transistor with multiple emitters. ROMs employing bipolar technology may have access times as low as 15 ns. The access time of a memory is the length of time that must elapse

after addressing before the addressed word is made available at the output of the memory.

The number of bits in a memory is defined to be equal to the product of the number of words and the number of bits per word. (In a ROM like that in Fig. 12.10-1a the number of diodes will generally be about one-half the number of bits, since we may expect 1s and 0s to occur with about equal frequency.) When the number of bits is large, say in excess of 1,000, MOS technology is preferred to bipolar technology since MOS devices are more economical of silicon die real estate. The implementation of a single-bit line (again, for ease of comparison, the W_3 output bit line) is shown in Fig. 12.10-1c. Assuming positive logic as usual and assuming NMOS, we see that W_3 is at logic 0 only when $Z_1 = 1$, as required by the truth table of Fig. 12.9-1a. Access times using MOS devices are typically about 400 ns.

12.11 PROGRAMMABLE AND ERASABLE ROMs

In a ROM, the address lines and the output word bit lines form a crossed array of lines, i.e., a grid structure. At each grid intersection is placed a device (diode, bipolar, or MOS transistor) or not, depending on whether the corresponding word bit is to be 1 or 0. (In cases where there is no special interest in the type of device, the coupling between address line and bit line is often shown simply by a dot at the grid intersection.) In a programmable ROM (PROM) the manufacturer locates a connecting device at every grid intersection. However, in series with each such device there is provided a fusible link. Any particular fusible link is located at the intersection of some line Z_i and some line W_j . By making connection to Z_i and W_j and passing an adequately large current through the link, the link can be burned out. Thus, the user of such a PROM may burn out links as necessary, leaving transistors only on locations required to establish the memory storage desired.

One type of erasable or alterable ROM uses floating-gate PMOS transistors. These are transistors in which at normal operating voltage the gate is entirely insulated and isolated from electrical connection to any other part of the integrated-circuit chip. It turns out to be possible to establish a negative charge on these gates by the application of a high voltage between source and drain. The negative charge left on the gate by such treatment leaves the corresponding transistor with a conducting channel. The ROM can be erased by exposure to ultraviolet light, which serves to discharge any charged gate.

12.12 APPLICATIONS OF ROMs

A ROM multiplier Consider that we want to perform the arithmetic operation of multiplication. As we have seen in Sec. 11.16, multiplication can be performed by a sequence of shifting operations, i.e., multiplying by powers of 2, and a sequence of additions. On the other hand, we may view a multiplication table

as a truth table. Thus, the entry in the multiplication table to effect the multiplication $11 \times 10 = 0110$, that is, $3 \times 2 = 6$, may be read to mean that a memory addressed by the input code $A_1 A_0 B_1 B_0 = 1110$ reads out the word $P_3 P_2 P_1 P_0 = 0110$.

When the multiplicands have many bits, straightforward multiplication by a ROM may get out of hand. For example, when the multiplicands consist of 8 bits, the product contains 16 bits and the number of bits which need to be stored in the memory is $16 \text{ bits/word} \times 2^{16} \text{ words} \approx 10^6$. A saving in bit storage can generally be accomplished by using a number of ROMs in a manner now to be illustrated.

Let us consider that we want to multiply two 4-bit numbers $A_3 A_2 A_1 A_0$ and $B_3 B_2 B_1 B_0$. We write

$$A_3 A_2 A_1 A_0 = A_3 A_2 00 + 00 A_1 A_0 \quad (12.12-1a)$$

$$\text{and} \quad B_3 B_2 B_1 B_0 = B_3 B_2 00 + 00 B_1 B_0 \quad (12.12-1b)$$

the product is then

$$P = (A_3 A_2 00)(B_3 B_2 00) + (00 A_1 A_0)(B_3 B_2 00) \\ + (A_3 A_2 00)(00 B_1 B_0) + (00 A_1 A_0)(00 B_1 B_0) \quad (12.12-2)$$

Note that the product P appears as the sum of four products. Four ROM multipliers are required, as indicated in Fig. 12.12-1. In these multiplications, however, the 0s in Eq. (12.12-1) are ignored. Thus, to effect the last product in Eq. (12.12-2) we use a ROM which accepts as input the address $A_1 A_0 B_1 B_0$ and yields one of 16 possible 4-bit output product words. We must now sum the outputs of the four multipliers, but in so doing we must take account of the differing numerical significance of the output bits of the different multipliers. As can be verified, the numerical significances are as given in the figure. Only one ROM yields bits of significance 2^1 and 2^0 . These are brought out and one ROM provides bits of significance 2^2 become bits of the final product. Three ROMs provide bits of significance 2^3 and 2^4 , and these are combined in two 2-bit adders. Similarly three ROMs provide bits of significance 2^5 and 2^6 , and these are combined in two other 2-bit adders. It is left as a student exercise (Prob. 12.12-2) to show that the disposition of carry outputs from the various adders is proper to assure a correct product result.

Each of the ROMs in Fig. 12.12-1 stores 16×4 bits. The four ROMs store $16 \times 4 \times 4$ bits = $2^8 = 256$ bits. If the multiplication had been effected by a single ROM, that ROM would need storage for 2^8 words \times 8 bits/word = $2^8 \times 8$ bits. Hence the present scheme reduces the required storage capacity by a factor of 8. The saving is more impressive as the number of bits increases. When the factors to be multiplied each have 8 bits, there is a saving by a factor of $2^7 = 128$. Of course, some of this saving is dissipated on account of the need to provide five adders. In addition, the use of adders, even with a look-ahead carry feature, results in a significant increase in the time required to perform the multiplication.

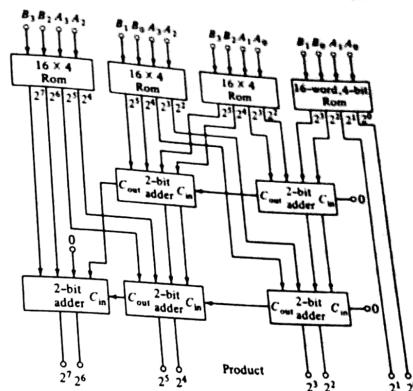


FIGURE 12.12-1
A ROM multiplier for two 4-bit numbers, illustrating a technique for reducing required storage capacity.

It is often required that a ROM multiplier produce an output having the same number of bits as each of the multiplicands. Thus, in the above example, where the inputs were each 4-bit words, the output would also be a 4-bit word. In this case, referring to Fig. 12.12-1, the output bits would consist of the four most significant bits. In this case it is possible to keep the maximum error between ± 8 by using additional circuitry (Prob. 12.12-1). If an error of 15 is permitted, the complexity of Fig. 12.12-1 can be reduced significantly to include only three ROMs and three adders.

The ROM as a look-up table ROMs are also widely used as look-up tables for mathematical functions such as logarithms, trigonometric functions, square roots, exponentials, etc. Techniques similar to that employed with the multiplier can be used effectively to reduce the storage capacity required of the memory. For example, consider a ROM for $\sin \theta$. If we wanted to provide for a resolution of 0.01° , then for the range $0 \leq \theta \leq 90^\circ$, a 9,000-word memory would be required. We may, however, write $\theta = I + F$, where I is the integral part of θ and F is the fractional part. Then

$$\sin \theta = \sin(I + F) = \sin I \cos F + \cos I \sin F \quad (12.12-3)$$

The implementation of Eq. (12.12-3) requires four smaller memories: two 90-word memories for $\sin I$ and $\cos I$ and two 100-word memories for $\cos F$ and $\sin F$. Further simplifications are possible in the matter of bits per word through the recognition that $\cos F \approx 1$ and $\sin F \approx F$.

12.13 BIPOLAR-JUNCTION-TRANSISTOR RANDOM-ACCESS MEMORY CELLS

The bit-storage cells in a random-access memory (RAM) may use BJT or may use MOS devices in either a static or dynamic mode. When high speed is required, the BJT is the RAM of choice inasmuch as it can be accessed in about 35 ns, in contrast to the MOS RAM, which has access times of approximately 400 ns. On the other hand, a BJT RAM is small, containing 1024 memory cells or less, while MOS RAMs are available with 4,096 memory cells.

In this section and in Sec. 12.14 we consider memory cells using the BJT. MOS cells are discussed beginning with Sec. 12.15.

The BJT RAM memory cell is simply a flip-flop. However some special features must be incorporated into it in order to facilitate addressing the cell, writing into it, and reading from it. Of course, given a flip-flop and the availability of gates as required, these three operations are easily effected. However, since useful RAMs may contain several hundred storage cells, these operations must be done with economy.

A RAM memory cell using multiple-emitter transistors is shown in Fig. 12.13-1. The cell itself is the flip-flop composed of T_1 and T_2 . The remaining circuitry in the figure provides a mechanism for writing and reading data. This auxiliary circuitry may service additional memory cells as well, which may be bridged across the data lines; hence the dashed extensions of the data lines. Commercially available bipolar memories of the type being discussed operate with a supply voltage in the range 3.5 to 5.0 V. We may then reasonably consider that logic 0 represents any voltage less than about 0.3 V and logic 1 represents any voltage above about 3.0 V. As indicated in the figure, the signals X and Y make excursions between these logic levels. These signals are used to address the cell. As we shall now see, the cell is accessed for reading or writing when X and Y are simultaneously at logic 1.

Let X and Y be at logic 0 (≤ 0.3 V), and let the read-write line be at logic 0 as well. Then T_3 and T_4 are ON, with collector voltages also at logic 0, and diodes D_1 and D_2 do not conduct. If, say, the state of the flip-flop is such that T_1 is conducting and T_2 is OFF, then emitter current will flow in E_x and E_y . A bias voltage of 0.5 V is applied through the resistor R_3 to emitter E_D . E_D is more positive than E_x and E_y by at least 0.2 V, and hence E_D does not conduct. The transistors T_5 and T_6 are also OFF, and the data output terminal is at logic 1 and will persist at this logic level independently of the state of the flip-flop. Now let the cell be addressed by raising both X and Y to logic 1. Then the currents through E_x and E_y will be diverted to E_D . A component of this current

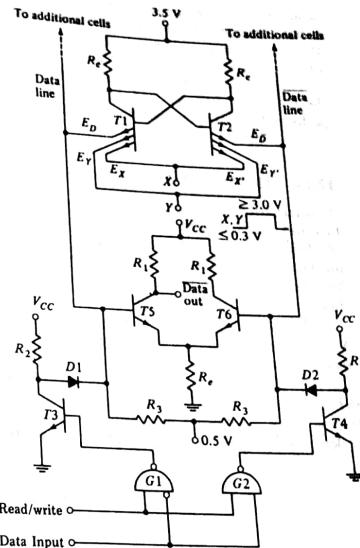


FIGURE 12.13-1
A RAM BJT memory cell. Provision is made for reading and writing.

will flow into the base of T_5 and the data out terminal will assume the same logic level present at the collector of T_1 . Thus, with the read-write line at logic 0 (in which case gates G_1 and G_2 are disabled against the entry of data) the operation of addressing the cell provides a reading of the cell.

Now, again, let $X = Y = 1$, so that the flip-flop is using emitters E_D and E_B , and suppose that the read-write line is at logic 1. Now gates G_1 and G_2 are enabled, and if the data line is at logic 1, T_3 will remain ON but T_4 will turn OFF. The collector of T_4 will rise, pulling with it the emitter E_D of T_2 . Hence, no matter what the original state of the flip-flop, it will be forced to the state in which T_2 does not conduct. Hence, the logic level at the collector of T_2 will become the logic level of the data input. If, however, the cell had not been addressed, E_D and E_B would not have been carrying current and the flip-flop

would not have responded to the writing operation. Of course, even if the cell is not addressed, the *data output* terminal will respond to the writing operation. Presumably, however, the memory is part of a system that "knows" that a writing operation is in process and will ignore the output until instructed to *read*.

The three-emitter transistor flip-flop cell of Fig. 12.13-1 is addressed when the logical product XY and $XY = 1$. When there is no need for the cell to perform this AND operation, one of the emitters may be omitted on each transistor.

12.14 OTHER BIPOLAR-TRANSISTOR MEMORY CELLS

An alternative bipolar memory cell is shown in Fig. 12.14-1. Schottky transistors (T_1 and T_2) and Schottky diodes (D_1 and D_2) are employed. In the standby condition (cell not addressed) the address line X is held at logic level 1, which for the present cell is about 2.5 V. At this level, the flip-flop remains operational. The data lines are biased at 1.6 V so that when the cell is not addressed, the diodes (D_1 and D_2) are back-biased.

The cell is addressed by lowering the address line to logic 0 (≈ 0.3 V). Assume, then, that before being addressed the cell was in the state in which T_1 was conducting and T_2 was OFF. When X drops to 0.3 V, the base of T_1 will drop to $0.3 + 0.75 = 1.05$ V. The forward voltage of a Schottky diode is about 0.45 V. Hence the data line will be pulled down slightly from 1.6 V to $1.05 + 0.45 = 1.5$ V. Allowing as usual, a voltage $V_{CE}(\text{sat}) = 0.2$ V across T_1 , the collector of T_1 will be pulled down to $0.3 + 0.2 = 0.5$ V. If we neglect the

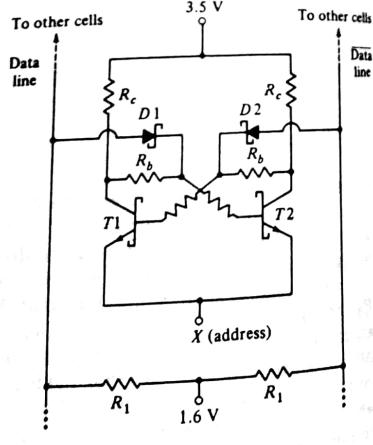


FIGURE 12.14-1 Bipolar-transistor memory cell

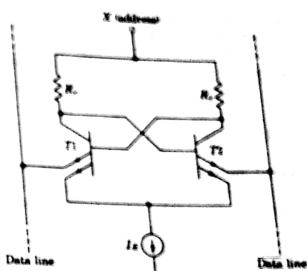


FIGURE 12.14-2
A bipolar-transistor memory cell which operates in the manner of emitter-coupled logic.

drop across the resistor R_b , we then find that the data line is pulled down rather substantially from 1.6 V to $0.5 + 0.45 = 0.95$ V. In practice, the resistor R_b is small enough ($\approx 1\text{ k}\Omega$) to ensure that even when the drop across R_b is taken into account, the change in voltage on the data line is still much larger than the voltage change on the data line. The difference in the voltage change between the data line and the data line is used to read the cell.

It is important to remember that the data and data lines go to many other cells. When the cell shown in the figure is addressed, the other cells are not and diodes D_1 and D_2 of these other cells are cut off, disconnecting the other cells from the data and data lines.

To write into the cell we first address it by setting X to 0.3 V. Assume again that initially T_1 is ON and that we want T_2 to be ON. Then we force the data line to go to 2.8 V. As a result T_2 will turn ON, and turning T_2 ON will turn T_1 OFF. The other cells, which have not been addressed, do not change state, for the voltage at the collector of each T_1 of these other cells will be $2.5 + 0.2 = 2.7$ V, and to make diode D_1 conduct it would have been necessary to raise the data line to $2.7 + 0.45 = 3.15$ V rather than 2.8 V.

The Schottky cell of Fig. 12.14-1 has some advantage compared with the multiple-emitter cell of Fig. 12.13-1. With the Schottky cell, the standby voltage across the cell is $3.5 - 2.5 = 1.0$ V. With the multiple-emitter cell this voltage is $3.5 - 0.3 = 3.2$ V. Hence, if the collector resistors R_c are the same in the two cases, the power dissipation in the Schottky cell will be appreciably smaller. Power dissipation can be reduced by increasing R_c . However, such increase in collector-circuit resistances would lower the speed. On the other hand, the Schottky cell does not have the two-address-line feature of the multiple-emitter cell.

Another memory cell is shown in Fig. 12.14-2. Here multiple-emitter transistors are again used. However, the addressing is done from the collector

side of the transistors so that the common emitters can be returned to a constant current source (possibly a large resistor). In this way the cell can be operated in the manner of emitter-coupled logic, where transistors T_1 and T_2 are not allowed to saturate.

12.15 MOS RAMS

A six-transistor static cell A six-transistor static MOS memory cell is shown in Fig. 12.15-1a. A bit is stored in the flip-flop, which consists of two cross-coupled inverters. Transistors T_1 and T_2 are the driver and load of one inverter, and T_3 and T_4 serve correspondingly for the other inverter.

The memory cell shown is addressed by setting X and Y to logic 1. Setting $X = 1$ connects the cell to the data line and the data line. To write into the cell we set $W = 1$. This connects the data input terminal to node D as T_5 , T_7 , and T_9 are ON. If the data input is at logic 1, this raises the gate of T_3 , turning it ON and making node $\bar{D} = 0$. If the data input is at logic 0, then T_3 would be turned OFF and \bar{D} would be at 1. To read the state of the flip-flop we set $R = 1$. This connects the data output terminal to \bar{D} since now T_6 , T_8 , and T_{10} are ON. Thus, the complement of the data level written into the cell is read.

In general, in a RAM, there are many memory cells connected to the same input and output lines. Such a configuration is shown in Fig. 12.15-1b. In this figure there are $\alpha\beta$ memory cells, each containing six transistors. Note that there are β columns of cells, each with a different Y address, and α rows of cells, each with a different X address. There is however, a single input transistor T_9 to connect the data input terminal to the selected memory cell when the write instruction is given and a single output transistor T_{10} to connect the selected memory cell to the data output terminal during the read operation.

A four-transistor dynamic cell In the dynamic memory cell of Fig. 12.15-2 the transistor count is reduced from six to four with a corresponding saving of real estate on the silicon chip and a saving of power. The cell is composed of transistors T_1 , T_2 , T_5 , and T_6 . Transistors T_7 and T_8 as well as transistors T_{11} and T_{12} serve all cells having the same column Y address. Transistors T_9 and T_{10} are common to all of the cells in the memory, as in Fig. 12.15-1b.

The state of the cell is stored on the stray capacitances C_1 and C_2 , whose presence is essential. These capacitors become accessible to the data terminals when the transmission gates T_7 and T_8 as well as T_5 and T_6 are all made to conduct by simultaneously raising the X and Y addresses to logic 1.

In one state of the cell, the voltage across C_1 is larger than the threshold voltage of T_1 , and T_1 is ON. Correspondingly C_2 has zero voltage, and T_2 is OFF. In the other state the voltages of C_1 and C_2 and the conducting states of T_1 and T_2 are reversed. The cell having been accessed, the state of the cell can be read by setting $R = 1$. We can write into the cell by setting $W = 1$.

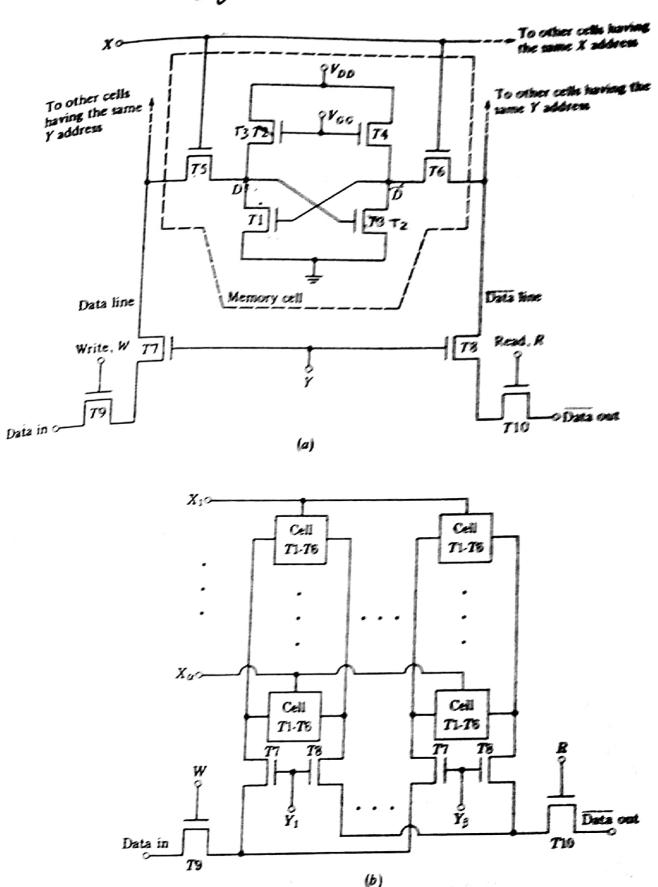
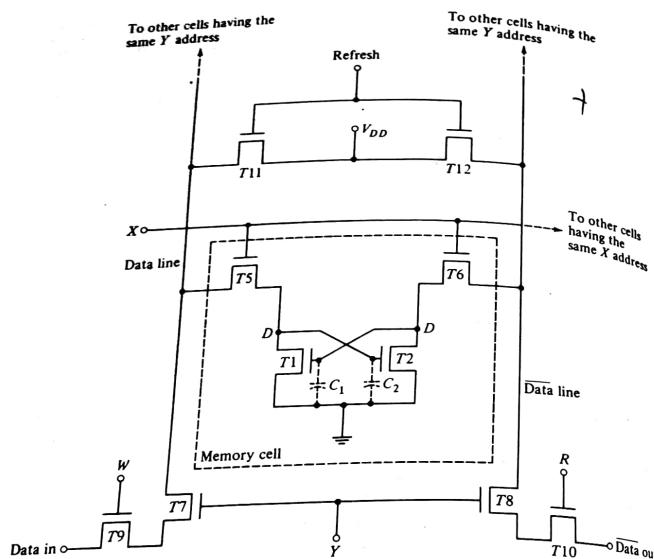


FIGURE 12.15-1
(a) A static MOS memory cell. (b) The interconnection of cells to form a RAM.

FIGURE 12.15-2
A four-transistor MOS dynamic cell.

If we have not performed a write operation for an extended time, then because of leakage of capacitor charge the information in the cell may be lost. It is therefore necessary to *refresh* the cell periodically. This refreshing operation is accomplished by allowing brief access from the supply voltage V_{DD} to the cell. This access becomes available when the X address and the refresh terminal are simultaneously at logic 1 so that T_5 and T_6 as well as T_{11} and T_{12} are thereby turned ON. Suppose now that initially T_1 is ON, T_2 is OFF, the voltage across C_1 is $V_{C_1} > V_T$, the threshold voltage, and $V_{C_2} = 0$ V. During the refresh interval V_{DD} is applied through T_{12} and T_6 to C_1 paralleled by T_2 . However, T_2 is OFF, and hence all the current from V_{DD} will be directed into C_1 , allowing C_1 to replenish any charge lost due to leakage. Similarly V_{DD} is applied to C_2 which is in parallel to T_1 . But T_1 is ON, and hence C_2 will not charge as rapidly as C_1 . Observe that during the refresh interval T_6 and T_{12} become a load for the driver transistor T_2 and that T_5 and T_{11} become a load for T_1 .

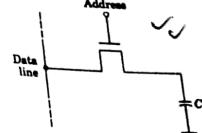
Hence, during this refresh interval the cell becomes a conventional flip-flop consisting of two cross-coupled inverters. In any event it is clear that whatever the initial state of the flip-flop, during the refresh interval this initial state is reinforced.

A three-transistor memory cell It is apparent that if a bit is to be preserved through the charge stored on capacitors (as in Fig. 12.15-2), then a single capacitor is enough. Such a rudimentary memory is indicated* in Fig. 12.15-3. Of course, additional circuitry will be required to allow reading, writing, and refreshing, but each component of this extra circuitry will be able to service many memory cells. We are concerned principally with the components which must be duplicated in every cell.

An important difficulty associated with the simple one-transistor cell of Fig. 12.15-3 is that during a read operation the storage capacitor C will discharge into the data line. Unless the storage capacitance is large in comparison with the data-line capacitance, the read operation may well be destructive. That is, the process of reading a bit may well erase the bit from the memory. Suppose, on the other hand, that C is made quite large. Then during a write operation a large capacitor needs to be charged, and the interval which needs to be devoted to a write operation will thereby be extended.

One resolution of the difficulty is presented in Fig. 12.15-4, where we have provided separate access paths to the capacitor C for writing and reading. However, now a three-transistor memory cell is required. In this cell, transistor T_1 is used during the write instruction, while transistors T_2 and T_3 are used when the logic state of capacitor C is to be read. Since the capacitor is the gate capacitor of T_2 , it is isolated from the output data line. A refresh circuit is also needed, of course, inasmuch as the charge on capacitor C can leak off through T_1 . Refreshing is accomplished using an inverter consisting of transmission gate T_9 and the inverter amplifier formed by capacitor C , and transistors T_{10} and T_{11} . Note the similarity between this refresh amplifier and the inverter shown in Fig. 12.3-4, consisting of T_1 , T_2 , T_3 , and C_1 .

To access the cell we set X and Y to 1. As before, there are many cells having the same Y address. To write into the cell we disconnect the refresh cells by setting $P = 0$. We then set $W = 1$. This connects the data-input circuit to setting $P = 0$.



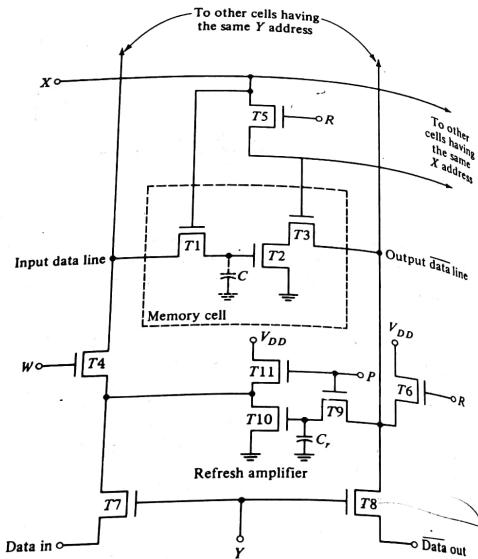


FIGURE 12.15-4
A three-transistor dynamic memory cell.

terminal across capacitor C since $T7$, $T4$, and $T1$ are ON. Capacitor C then charges to the state of the data input. To read the cell we set $R = 1$ ($W = 0$), which turns $T6$ and $T3$ of the addressed cell ON. The data-output line is then connected to the drain of $T2$ since $T3$ and $T8$ are ON. The complement of the logic level stored on C is read at the data-output terminal. Note that transistor $T6$ acts as a load for $T2$ during the read operation so that $T2$ and $T6$ form an inverting amplifier.

To refresh the cell we set $Y = 0$, $X = 1$, $P = 1$, and $R = 1$. The data-input and the data-output terminals are now disconnected from all the cells in the memory. The complement of the logic level of capacitor C is now transferred through $T9$, and this level is stored on capacitor C_r . The input terminal P , which permits the connection of C to the output data line, is called the precharge input. This term is used because when $P = 1$, C precharges to the complement

of the level on C . After this precharging is accomplished, R is set to 0 and W is then set to 1. The output of the inverter, $T10$ and $T11$, then refreshes the charge on capacitor C . Note that initially the refresh-amplifier output is disconnected from $T11$ and capacitor C . This is to ensure that $T11$ does not initially discharge capacitor C until after C has been precharged to the correct level. If this precaution were not taken, capacitor C might be discharged erroneously.

The memory cells having the same Y address are often refreshed sequentially, the cell having address X_1 being refreshed first, then the cell with address X_2 , etc. For each column of cells there is a single refresh circuit; thus if there are β columns in the RAM (see Fig. 12.15-1b), there are β refresh circuits. The β cells having the same X address are refreshed simultaneously. Thus, if there are α rows, and if it takes a time T_r to refresh a single cell, the entire memory is refreshed in the time αT_r .

Note that although there are only three transistors in the memory cell shown in Fig. 12.15-4, many auxiliary transistors are needed to implement the reading, writing, and refreshing operations. However, these transistors are shared by other memory cells.

12.16 ORGANIZATION OF A RAM

The organization of a memory with storage facility for M words each of 3 bits is shown in Fig. 12.16-1. Cell $(1, 1)$ stores the first bit of the first word, cell $(1, 2)$ the second bit and so on. These first word-storage cells are addressed by raising the level of line Z_1 . If the memory stores M words, M address lines Z_1, Z_2, \dots, Z_M are required. The memory-storage location is presented here in coded form through the λ address bits $A_0 A_1 \dots A_{\lambda-1}$ (where $M = 2^\lambda$). As required, then, a decoder has been interposed between the memory cells and the coded input address. As discussed in Sec. 12.9, such a decoder singles out one and only one of the address lines, the line so selected depending on the input address. Not all commercially available memories provide such decoding, and in such cases the decoders must be provided by the user. When a decoder is already part of a memory unit, the manufacturer will characterize his product as being "decoded" or "fully decoded."

All the cells $(C_{1,1}, C_{2,1}, \dots, C_{M,1})$ intended to store first bits of words are connected to a common pair of data lines. All second-bit cells and all third-bit cells are similarly connected. The block marked I/O (input-output) represents all the circuitry shown in Fig. 12.13-1 or 12.15-4 with the exception of the memory-cell flip-flop itself. The bits b_{11}, b_{12} , and b_{13} are input bits which will be written into the memory when the read/write (RW) line goes to logic 1. The bits b_{01}, b_{02} , and b_{03} are the output bits which will be read out of the memory when $RW = 0$.

If the memory cell has two input address terminals, as in Figs. 12.13-1 and 12.15-4, we can arrange to make each cell separately and individually accessible. In such a case a memory with M cells can be used to provide

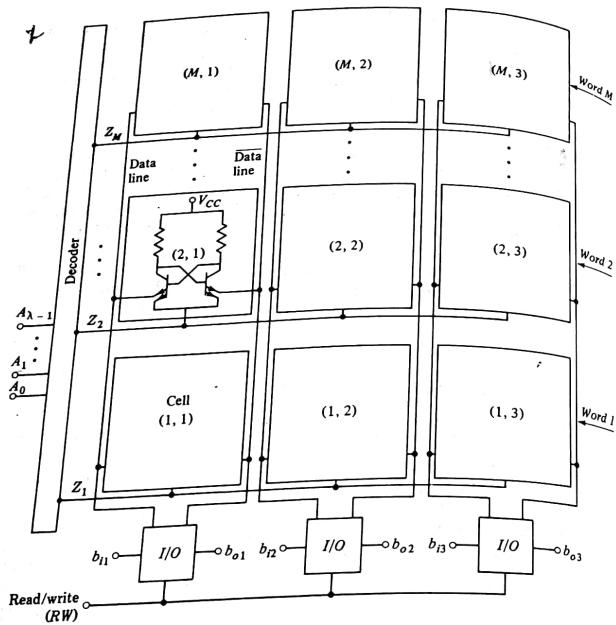


FIGURE 12.16-1
The organization of a M -word 3-bit/word memory.

storage for M words, each word having just 1 bit. The organization of such a memory is shown in Fig. 12.16-2. Note particularly that two decoders are required. At any time only one output line of the X decoder will be selected and only one line of the Y decoder will be selected. Correspondingly only that cell will be selected (addressed) which is at the intersection of these two selected lines. All the data lines of all the cells are paralleled, and only a single input-output stage is required to handle the 1-bit word.

This present arrangement allows the use of simpler decoders since some of the decoding is done in the memory cell itself. For example, suppose we need a memory with 256 (2^8) words each of 1 bit. If the cells had single address lines, we would need a decoder with 256 output lines. Suppose, however, the

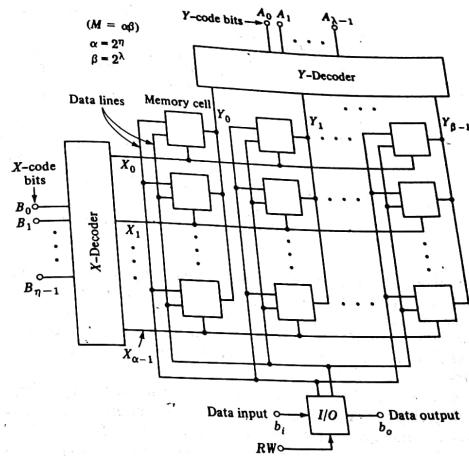


FIGURE 12.16-2
Organization of a memory with N 1-bit words.

cells were arranged, as in Fig. 12.16-2, in a square array ($\alpha = \beta = 16$). Then we should require two decoders, but each decoder would have only 4 input lines and 16 output lines.

If the memory cells employed in the RAM shown in Fig. 12.16-2 are dynamic, each column of cells has its own refresh circuit, as in Fig. 12.15-4. Thus, in a 1,024-bit memory containing 32 columns, 32 refresh circuits are used.

12.17 PARALLELING OF SEMICONDUCTOR MEMORY INTEGRATED-CIRCUIT CHIPS

On an integrated-circuit memory chip, the number of bits in a word is equal to the number of memory cells that have a common address line, and the number of words is equal to the number of separately addressable groups of bits. Manufacturers generally specify word and bits per word storage capacity through

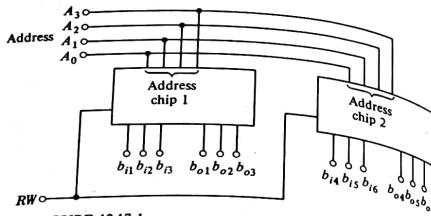


FIGURE 12.17-1
Paralleling memory chips to increase the number of bits per word.

the term *organization*. Thus, typically, a small memory may be characterized as having a capacity of 64 bits organized in 16 words of 4 bits each. (The number of words is invariably a power of 2 in order to put to effective use all the outputs of a decoder from a binary-code input.) As noted, the chip may also have a decoder and will have, as well, the required input-output circuitry.

The capacity of a memory system can be increased by paralleling chips. Paralleling chips in a manner which holds the number of words fixed but increases the number of bits per word is shown in Fig. 12.17-1. Here we represent the case of two identical chips, each of 48-bit capacity, organized in 16 words with 3 bits/word. The words are addressed by a 4-bit address code A_3, A_2, A_1, A_0 . The address inputs of the two chips are paralleled, as is the read/write (RW) input. The 3 output bits of chip 1 (b_{01}, b_{02}, b_{03}) and the 3 output bits of chip 2 (b_{04}, b_{05}, b_{06}) become the 6 output bits (b_{01}, \dots, b_{06}) of the new enlarged word. Similarly the input bits of the 6-bit word are $b_{11} \dots b_{16}$. Additional chips, of course, may be paralleled to increase the number of bits per word further.

If the 1-bit word memories of the type shown in Fig. 12.16-2 are paralleled, the number of bits in a final output word is directly equal to the number of cells paralleled. The X -code input bits are applied to the X decoders of each chip, and the Y -code input bits to the Y decoders of each chip. This paralleling is illustrated in Fig. 12.17-2. Here we have drawn all the A and B address lines as single lines to simplify the figure.

Expansion of 1-bit per word memory chips Consider again that we have memory chips, each with M bits, in which each individual bit is separately addressable. We have already seen, as in Fig. 12.17-2, how N such chips may be organized into a memory of M words each of N bits. Suppose, however, that while N bits per word are adequate, we require more than M words. Let us say that we require $L \times M$ words. Such an expanded memory may be constructed by paralleling L groups of N chips, as shown in Fig. 12.17-3.

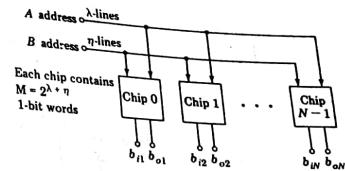


FIGURE 12.17-2
Paralleling chips to increase the number of bits per word. The special case in which each chip provides only one bit.

The chips in the top row store the first M words, the next M words are stored in the second row, and finally the L th group of M words is stored in the bottom row. In order that it be possible to select a word in a particular row, the manufacturer makes available on the chip an additional address line called *chip select*. When the chip-select line is $CS = 0$, the chip decoder is disabled and no word on the chip is addressed. When the line is at $CS = 1$, the chip decoder is enabled and words are accessed in normal fashion.

In Fig. 12.17-3 the arrow marked A at each chip represents all the address

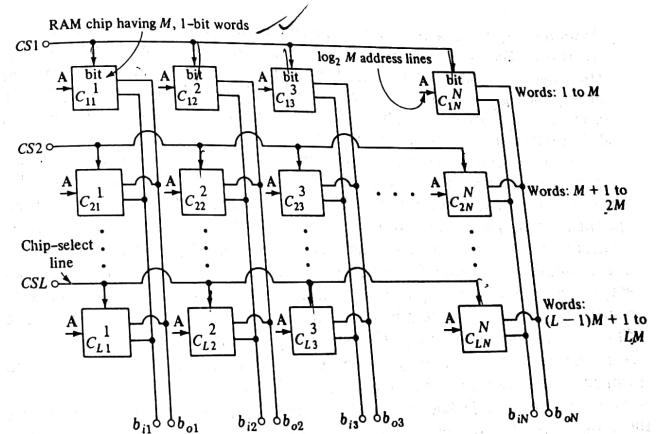


FIGURE 12.17-3
Connection of $M \times 1$ RAM chips to form an LM -word by N -bit memory.

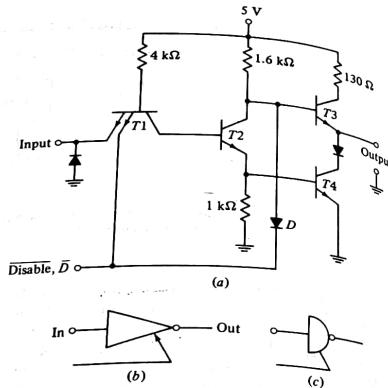


FIGURE 12.17-4
(a) A TTL tristate output stage; (b) and (c) circuit symbols for the stage.

input lines by which a single bit on the chip itself is addressed. The total input address applied to the memory, consisting of the address input lines A and the chip-select line CS , singles out a single chip-select line and the word is read from the corresponding row of chips. If we allow ourselves to compare reading from a memory with reading from a book, the chip-select selects a page and the address A then selects a word on the page.

Next we must take into account that when paralleling chips to increase the number of words it is necessary to connect in parallel the output bit lines of each of the other chips. Thus, for example, the output line of the parallel arrangement which bears the first bit of the word must be connected to the first-bit output terminals of each of the other chips; that is, $C_{11}, C_{21}, \dots, C_{L1}$ are in parallel. It is therefore necessary to devise for a memory chip an output stage which allows the output logic level of a selected chip to be unaffected by the connection of the output terminals of other chips which are not selected. The connection of the output terminals of other chips for this purpose. This stage is, of course, readily recognized as the standard TTL gate modified through the addition of a disabling input.

When the disable line is at logic 1, diode D does not conduct nor does the emitter of $T1$, which is connected to this line. Hence with $\bar{D}=1$ a straightforward output stage is provided with all of the usual advantages of

active pull-up (see Chap. 5). However when $\bar{D}=0$, both $T3$ and $T4$ will be off and the output terminal will be entirely isolated. The impedance seen looking back into the output terminal will be very high, being less than infinite only because of leakage and stray capacitance. The stage of Fig. 12.17-4 is often described as a tristate stage, the available states being logic 0, logic 1, and the third state, a nominally isolated high-impedance state. Symbols for the tristate stage are shown in Fig. 12.17-4b and c.

When memory chips are paralleled for the sake of increasing the number of words, the tristate output stage is enabled and disabled by the chip-select input. Only the output stages on the selected chip are enabled. CMOS gates are also available with a tristate output.

12.18 THE CHARGED-COUPLED DEVICE (CCD)

We have seen that an array of MOSFET devices fabricated on a silicon chip and forming an integrated-circuit dynamic-shift register may serve as a sequential memory. We consider now a different MOS dynamic-shift register sequential memory using the principle of the charge-coupled device (CCD). While the CCD memory will operate at about the same speed as the MOSFET memory, the CCD memory will dissipate appreciably less power. However, the most important relative advantage of the CCD memory is that it may be fabricated with a density of bits which is of the order of three times the density feasible with MOSFET memories. As we have noted, there is an advantage to using silicon chip "real estate" as economically as possible, since, as the area of the chip increases, there is a corresponding increase of the likelihood of encountering an imperfection in the chip. Hence the improved density possible with CCD devices results in better yields and lower cost. A still further advantage of the CCD device is that it requires a much simpler fabrication procedure involving many fewer operations than is required in MOSFET and in bipolar technology.

The structure of the CCD is represented in Fig. 12.18-1. The device is fabricated on a semiconductor substrate. We have indicated n-type silicon for the sake of being specific but p-type would serve as well. The substrate is covered with an insulating layer of silicon dioxide and on the oxide an array of closely spaced metal electrodes is arranged. The metal-oxide semiconductor sandwich is reminiscent of the MOSFET structure. Note however, the absence of p-type regions which in the MOSFET serve as source and drain.

Let us now consider initially that, while the bottom of the substrate is maintained at ground (0 V), all the metal electrodes are maintained at the same fixed negative voltage, $-V$. As a matter of practice, the spacing between electrodes is extremely small in comparison with their width. (The spacing in Fig. 12.18-1 is grossly exaggerated). Hence, in effect, the metalized layer can be considered to be an equipotential surface at the voltage $-V$. Since the base of the substrate is an equipotential surface at 0 V, the equipotential surfaces generally are planes parallel to the faces of the structure.

If the voltage $-V$ on the metalization is large enough in magnitude so

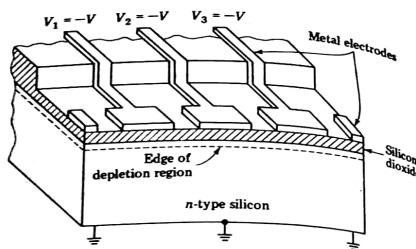


FIGURE 12.18-1
The structure of a charged-coupled device.

that it exceeds the threshold voltage V_T of the substrate, then there will form, under the oxide and adjacent to it, a depletion region. The depletion region is a region from which mobile carriers have been removed. In the present case, the relatively few minority holes present within the substrate will be drawn to the surface of the substrate under the oxide, while the majority carrier—electrons—will be pushed away from the surface. Since predominantly negative charge has been removed from the initially neutral region, the depletion region is left with a positive charge. Lines of electric field which originates from the negative charge on the metal will terminate on this positive charge. The boundary of the depletion region is established when enough positive charge has been provided to allow for the termination of all the electric-field lines. Like a typical equipotential surface, the depletion region boundary is a plane parallel to the structure surfaces (see Fig. 12.18-1).

Initially, the depletion region shown in Fig. 12.18-1 is devoid of mobile charges. However, as time progresses, holes present in the *n*-type substrate will diffuse into the depletion region. After a sufficient number of holes have entered the depletion region, there will be formed a conducting layer under the oxide precisely in the manner in which a channel is formed in an MOS device.

Depending on the nature and quantities of impurities in the semiconductor, the absence or presence of defects in the crystal structure of the semiconductor and other factors, the time that elapses between initial depletion and eventual channel generation will be in the range from tenths of seconds to tens of seconds or even longer. In any event, the time is long in comparison with the usual clock interval encountered in a digital system. This feature may seem surprising in view of our experience in connection with MOSFETs. There we noted that, when an appropriate voltage was applied to the gate, a channel formed immediately, i.e., in a matter of tens of nanoseconds. The difference arises from the fact that in the MOSFET, unlike the CCD device, there, there is a source of

minority carriers immediately to the side of the depleted region. Thus in the MOSFET we have, at the outset, a voltage applied between the source and drain. As soon as a voltage is applied to the gate to generate a depleted region, minority carriers from the source and from the drain rush into the depleted region and convert it instead to a conducting channel.

12.19 STORAGE OF CHARGE

Consider now the situation represented in Fig. 12.19-1a. Here the voltage on one of the metal electrodes has been made substantially more negative than the voltage on its neighbors; that is, there is a larger negative voltage on electrode 2 than on the others. Accordingly, the depletion region under electrode 2 extends more deeply than under the adjacent electrodes. The boundary of the depleted region now has the form indicated by the dashed line in Fig. 12.19-1a.

Qualitative and approximate plots of electrical potential as a function of x , the distance measured horizontally through the substrate, at a number of levels are as shown in Fig. 12.19b. At any level which passes through the depletion region the potential, under electrode 2, is depressed relative to the potential under its neighboring electrodes. The potential depression is more pronounced for a level AA' which is located closer to the electrodes and is less pronounced for a level CC' located further from the electrodes. We have taken account, in drawing

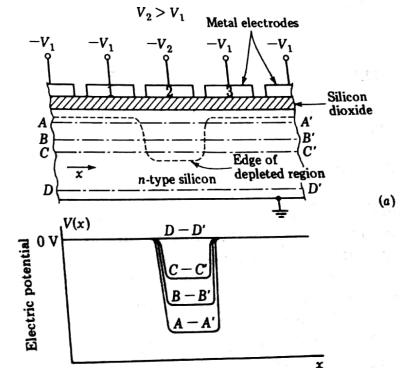


FIGURE 12.19-1
(a) The form of the depletion region when one electrode is made more negative than its neighbors. (b) The electric potential $V(x)$ through the substrate at various levels.

these plots, that outside the depletion region the substrate is a good conductor, being copiously supplied with majority carriers. Since no current flows, there is no electric field and, hence, no potential difference along the x axis within the n -type substrate.

The reader is now asked to recall that a charged particle of charge q in a field described by a (one-dimensional) potential $V(x)$ has exerted on it a force in the positive x direction given by $f = -q dV/dx$. Thus, in Fig. 12.19-1b where the potential is flat ($dV/dx = 0$), there is no force. In the region between electrodes 1 and 2 where dV/dx is negative, the force is to the right, and in the region between electrodes 2 and 3, where dV/dx is positive, the force is to the left. Thus, a positive charge introduced into the depletion region at any level is free to move about only within the "potential well" shown in b.

Finally, then, it appears that the application to a particular electrode of a negative voltage larger in magnitude than is applied to its neighbors generates under that particular electrode an extended depletion region which penetrates more deeply into the substrate. If now we introduce into this region some positive charge (in a manner yet to be described), this charge will be trapped in position. Eventually, this introduced charge will lose its identity because of the diffusion of new carriers into the depletion region as described above. But as long as the charge is identifiable, the presence or absence of such a charge may be used to represent the two logic levels. As we shall now see, there are means by which such trapped charges may be transferred from a position under one electrode to a position under the next, and so on, the whole structure then serving as a shift register.

12.20 TRANSFER OF CHARGE

A mechanism by which charge may be transferred laterally in the CCD is illustrated in Fig. 12.20-1. The CCD device itself, extending laterally in the x direction, appears in Fig. 12.20-1b. We consider initially at time $t = t_0$ that the voltage on electrode k is at $V = -V_2$ while the voltage on all other electrodes is $V = -V_1$. Then there is an extended depletion region under electrode k and a minimal depletion region under all other electrodes. Then, at time $t = t_0$, the voltage $V(x)$ across the substrate, at any level that intersects the extended depletion region, will appear as shown in Fig. 12.20-1c. Here we have assumed that we have devised, by some means, to introduce some positive charge into the depletion region. Somewhat fancifully we have represented the charge as sitting at the bottom of the potential well.

Now, as is indicated in Fig. 12.20-1d, let us arrange that at time $t = t_1$ the voltage on electrode $k + 1$ should also drop to $V = -V_2$. Then the potential profile takes on the appearance shown in Fig. 12.20-1d. The potential barrier to the right of the charge in Fig. 12.20-1c has now been removed. Accordingly, as indicated by the arrow in Fig. 12.20-1d, the charge will diffuse to the right. No charge transport takes place, at least initially, in the reverse direction simply

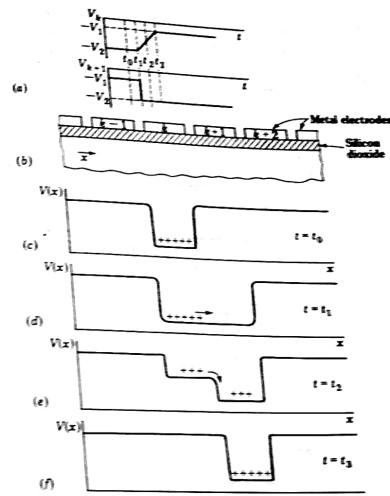


FIGURE 12.20-1
The mechanism of charge transfer. (a) The waveform V_k and V_{k+1} . (b) The CCD device. (c) through (f) The potential distributions at times t_0 through t_3 as identified in (a). The transfer of the charge is shown.

because there is no charge in the $k + 1$ region. If the voltage profile persisted as shown in Fig. 12.20-1d, then eventually the original charge would distribute itself evenly over the k and $k + 1$ region. However, we arrange, as shown by the waveforms of V_k and V_{k+1} in Fig. 12.20-1a, that the voltage V_k should now begin to return, relatively slowly, to the value $V_k = -V_1$.

As V_k rises, charge in the k region continues to move to the $k + 1$ region. However, now the mechanism of charge flow is due to a combination of diffusion and the presence of the electric field produced by the potential difference between the regions. Finally at $t = t_3$ and thereafter, as indicated in Fig. 12.20-1f, the charge is in the $k + 1$ region, having been transferred there from the k region.

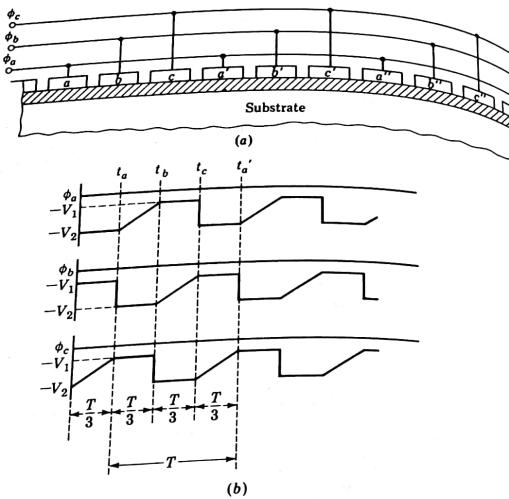


FIGURE 12.20-2
(a) The connection of a three-phase clocking waveform to the electrodes of a CCD device. (b) The waveforms of the clocking waveform phases.

During the interval from $t = t_1$ to $t = t_2$, while charge is being transferred from the k to the $k + 1$ region, it is necessary that the voltage at electrodes $k - 1$ and $k + 2$ be held at $V = -V_1$ so that no additional depletion region form under these electrodes. If a depletion region formed under electrode $k - 1$, some of the charge in the k region would be transported in the wrong direction. If of the charge in the $k + 2$ region would be transported in the wrong direction. If of the charge in the $k + 2$ region would be transported in the wrong direction. If of the charge in the $k + 2$ region would be transported in the wrong direction. If of the charge in the $k + 2$ region would be transported in the wrong direction. If of the charge in the $k + 2$ region would be transported in the wrong direction.

Taking account of the charge-transfer mechanism described, and taking account of the need to ensure that charge is not transferred backward or forward more than one step at a time, we may now consider how to effect a sequence of forward steps in the manner of a shift register. As shown in Fig. 12.20-2a every third electrode is connected to a common clocking bus and the sequence of forward steps in the manner of a shift register. As shown in Fig. 12.20-2b. The composite clocking waveform has a period T . Each of the parts consists of three subintervals of duration $T/3$. In these subintervals the waveform is at $-V_1$, or at $-V_2$, or is rising from $-V_2$ to $-V_1$. At the time

t_a , ϕ_b has been at $-V_1$ for a full subinterval $T/3$ so that there is certainly no charge under the b electrodes (b , b' , b'' , etc.). Also ϕ_c has just completed its rise to $-V_1$, as a result of which any charge which might have been under the c electrodes has been completely dumped into the regions under the a electrodes. Accordingly, at $t = t_a$, any charge (or a lack of charge, depending on whether a logic 1 or a logic 0 is being represented) is under the a electrodes. Consider specifically that there is such a charge under the electrode a in Fig. 12.20-2a. Then, during the time interval from t_a to t_b , this charge will be transferred to the region under the b electrode, the process of transfer being completed at time $t = t_b$. At time $t = t_b$, the charge will have been transferred to the region under electrode c . Finally, at time $t = t'_a$, after a total time T the charge which was under a at $t = t_a$ will now be found under a' . Observe, most importantly, that, when the cycle has been completed, charge has been transferred from one region to a second region *three electrodes away*. Thus it takes *three electrodes* to provide for storage and transfer of *one bit*.

We have described a CCD device in which the geometry of the electrodes and insulating oxide layer are such that a three-phase clock is required. This geometry has the difficulty, in the matter of fabrication, that the separation between electrodes is required to be inconveniently small. Other more complicated geometries have been developed which require a four-phase clock, a two-phase clock, and even a single-phase clock.

12.21 INPUT AND OUTPUT ARRANGEMENT

We have seen how a charge trapped in the depletion region under an electrode may be transferred laterally in shift-register fashion. We need to consider briefly how a charge is introduced at the input side and finally detected at the output side.

One method suitable for use in shift-register application is shown in Fig. 12.21-1a. Here, under an opening in the insulating oxide layer we have diffused a *p*-type region in the *n*-type substrate. We have also added an additional gating electrode preceding the clocked transfer electrodes. Consider then that the clock phase is such that there is a depletion region under the electrode next to the gating electrode as shown. Then the similarity to the situation that prevails in a MOSFET is rather apparent. If at this time a negative gating voltage is applied, a depletion channel will open under the gate and minority *p*-type carriers will flow across the channel from the *p*-type region to the depletion region. When the depletion region has its required complement of charge, the gating voltage is removed. Thereafter the succeeding phases of the clocking waveform transfer this charge laterally, leaving this region under the first transfer electrode available for the next charge.

Correspondingly, the packets of positive charge may be detected, at the end of the register, in the manner indicated in Fig. 12.21-1b. Again a *p*-type region has been diffused under an opening in the oxide insulation. An external voltage

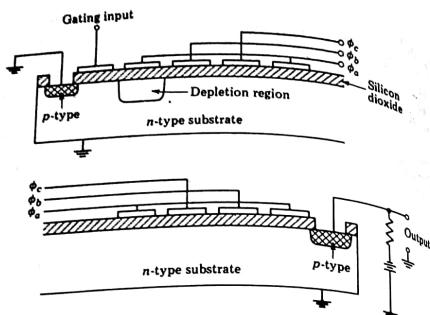


FIGURE 12.21-1
(a) A method of injecting a charge into a CCD shift register. (b) A method of detecting a charge at the end of a CCD shift register.

source reverse-biases the junction formed between the *n*-type substrate and the *p*-type diffusion. The minority carriers are transferred across this junction just as in the manner in which minority charges are collected by the collector junction of a conventional transistor. An output voltage is developed across the resistor in series with the reverse-biasing voltage.

The methods described for injecting and collecting charge have the disadvantage, in the matter of fabrication, that high-temperature diffusions into the substrate are required. Other methods that require no such diffusions are also available.

Finally, we must note that the process of charge transfer is not completely efficient. At each transfer of charge, from under one electrode to the next, some charge remains behind. In a long register it is accordingly necessary periodically to provide for refreshing. Refresh amplifier stages are fabricated directly on the substrate of the CCD without involving an inordinate area of the chip.

An example of a CCD memory Typical of the CCD memories presently commercially available is the Intel type 2416 memory. This memory is organized as 64 recirculating shift registers each of 256 bits. We have noted that because of the gradual disappearance of depletion regions a bit cannot be stored in one position indefinitely. There is, accordingly, a minimum time between shifts which must be observed. In the type 2416 this minimum time is 9 μ s. The maximum shift rate is 2 MHz. A four-phase clocking waveform is used. Average power dissipation is about 20 μ W per bit.

REFERENCES

- 1 Seurin, C. H., and M. F. Tompsett: "Charge Transfer Devices," Academic Press, 1975.
- 2 Boyle, W. S., and G. E. Smith: Charge-Coupled Semiconductor Devices, Bell System Tech. J., vol. 49, April 1970, pp. 587-600.
- 3 Boyle, W. S., and G. E. Smith: Charge-Coupled Devices—A New Approach to MIS Device Structures, IEEE Spectrum, July 1971, pp. 18-27.
- 4 Altman, L.: The New Concept for Memory and Imaging: Charge Coupling, Electronics, June 21, 1971, pp. 50-59.
- 5 Kosonocky, W. F., and J. E. Carnes: Charge-Coupled Digital Circuits, IEEE J. Solid State Circuits, vol. SC-6, no. 5, October 1971, pp. 314-322.

14

ANALOG-TO-DIGITAL CONVERSATIONS

ANALOG-TO-DIGITAL CONVERSATIONS 487

analog signal $M_i(t)$, the processing may consist of some digital operation such that when an output analog signal $M_o(t)$ is eventually reconstituted, the signals $M_i(t)$ and $M_o(t)$ are related by the transfer function of a *digital filter*.

At the input end of such a digital processing system, the overall process of converting an analog signal to a digital form involves a sequence of four individual processes, called *sampling*, *holding*, *quantizing*, and *encoding*. These processes are not necessarily performed as separate operations. Rather generally, as discussed in Chap. 13, sampling and holding are done simultaneously in a type of circuit referred to as a sample-and-hold (S/H) circuit, while quantizing and encoding are done simultaneously in a circuit referred to as an *analog-to-digital (A/D) converter*. After the digital processing is completed, the reconstitution of an analog output signal is accomplished by the operations of *digital-to-analog (D/A) conversion* followed by *filtering* or *smoothing*. In this chapter we shall discuss the circuits used to perform each of these functions.

14.2 THE SAMPLING THEOREM¹

The validity of the entire process depends fundamentally on the *sampling theorem*, which we shall now state. This theorem is so well known and is discussed at such lengths in texts on communication theory and systems analysis that we shall not prove it here.

Let $M(t)$ be a signal which is band-limited such that its highest-frequency spectral component is f_m . Let the values of $M(t)$ be determined at regular intervals $T_s \leq 1/2f_m$; that is, the signal is to be sampled regularly every T_s , or more frequently. Then these samples uniquely determine the signal, and the signal can be reconstructed from these samples with no error. The signal $M(t)$ can be reconstructed precisely by transmitting the samples through an ideal low-pass filter which has a flat response at least to f_m and cuts off at a frequency less than or equal to $f_s - f_m$, where $f_s = 1/T_s$.

The time T_s is called the *sampling time*. Note that the theorem requires that the *sampling rate* f_s be rapid enough for at least two samples to be taken during the course of the period corresponding to the highest-frequency spectral component of the signal $M(t)$.

The significance of the sampling theorem is illustrated in Fig. 14.2-1. In Fig. 14.2-1a the sampling of the band-limited signal is accomplished by the FET switch and its associated sampling signal. The sampling signal, which is the FET gate-control signal, consists, as shown, of a pulse sequence, the pulses having a duration τ and a period T_s . The gate allows transmission only during the interval τ . The signal, except for a multiplicative constant α is reconstituted at the output of the low-pass filter. A portion of a typical signal is shown in Fig. 14.2-1b, and the sampled waveform is shown in Fig. 14.2-1c. The sampled waveform $M_s(t)$ consists of pulses. The duration of each pulse is equal to the time interval during which the FET switch is closed. During each sampling

14.1 INTRODUCTION

Digital signals, ideally at least, are represented by waveforms which make abrupt transitions between two values. Signals which may assume any value in a continuous range are called analog signals. When analog signals must be processed, there is often a great advantage in converting the signal to digital form so that the processing can be done digitally.

By way of example, an analog voltage may be a fixed (dc) voltage, and the required processing may consist of determining its value. An analog voltmeter will display the voltage value through the position of a pointer on a scale panel. Or the analog signal may be the time-varying output voltage $M(t)$ of a microphone. The required processing may then consist of transmitting $M(t)$ to the input of a distantly located loudspeaker so as to minimize the effect of noise (random unpredictable disturbance) which is invariably superimposed on the signal during transmission. A most effective way of suppressing noise is to transmit the signal digitally. A communication system which operates in this way, i.e., converts the analog signal to digital form and then reconstitutes the analog signal, is called a *pulse-code modulation system*. Or, again, given an input

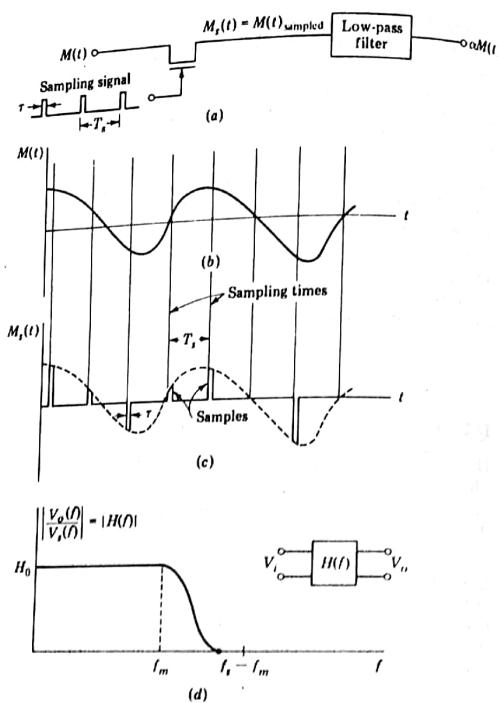


FIGURE 14.2-1
 (a) A signal $M(t)$ is sampled and reconstructed. (b) A signal $M(t)$. (c) The signal $M(t)$ sampled. (d) The transfer characteristic required of the filter.

interval $M_s(t) = M(t)$; that is, the top of each pulse follows the contour of $M(t)$. Outside the sampling interval $M_s(t) = 0$.

The characteristic required of the filter, for $M(t)$ to be reconstituted without error, is shown in Fig. 14.2-1d. Here $H(f)$ is the transfer function of the filter; that is, $H(f) = V_o(f)/V_s(f)$. We note that $H(f)$ must be flat at least up to the frequency f_m , which is the highest-frequency component of $M(t)$, and $H(f)$ must fall to zero before the frequency $f_s - f_m$. By way of example, say $f_m = 1$ kHz; then the minimum allowable sampling frequency is $f_s = 2f_m = 2$ kHz. To allow some latitude let us take $f_s = 2.5$ kHz. Then $f_s - f_m = 2.5 - 1.0 = 1.5$ kHz. It would then be required that $|H(f)|$ be constant, $|H(f)| = H_0$, from 0 to at

least 1 kHz and then fall to zero at a frequency ≤ 1.5 kHz. If we had selected $f_s = 2f_m = 2$ kHz, we would have had $f_s - f_m = 2f_m - f_m = f_m$. In this case we would require that $|H(f)| = H_0$ up to f_m and then $|H(f)|$ should drop abruptly to zero.

While we have made no explicit reference to the phase characteristic of the filter, it should be noted that in the filter passband (0 to f_m) the phase characteristic must be linear. Such is the requirement in the passband of any filter if the filter is to pass, without distortion, any signal whose spectral components all lie in the filter's passband.

Suppose, referring to Fig. 14.2-1c, that we arrange that $\tau = T_s$, that is, that the sampling interval be as long as the interval between samples. In this case the FET switch will transmit all the time, and the amplitude of the reconstituted signal $M_o(t)$ is $M(t)$, that is, all of the signal. More generally, the FET switch transmits for only a fraction τ/T_s of the time. It seems reasonable, then, that in this situation the fraction of the original signal $M(t)$ available in the reconstituted signal should be $(\tau/T_s)M(t)$. The same result can be seen in another way. A low-pass filter is a time-averaging device. Thus the extent to which $M(t)$ is available in the sampled signal should be proportional to the width τ of the sample pulses in the sampled signal, the available signal being $M(t)$ itself when $\tau = T_s$. Finally, taking account of the gain H_0 of the low-pass filter itself, we would expect the recovered output signal to be

$$M_o(t) = \frac{\tau}{T_s} H_0 M(t) \quad (14.2-1)$$

In Fig. 14.2-1a the distance between the signal source $M(t)$ and its eventual destination might be so large that it would be appropriate to refer to the connection between source and destination as a *communication channel*. The channel connection might be composed of wires, or it might consist of a radio link. The physical facility which constitutes the communication channel might well have been quite expensive to establish, and we would be interested in making maximum possible use of it. We note that in the transmission of the signal $M(t)$ the channel appears to be "in use" only a fraction τ/T_s of the available time, and we might inquire whether when the channel is not being used for $M(t)$ it might be used to transmit other signals. Such simultaneous transmission of signals is indeed possible and is referred to as *time-division multiplexing*.

14.3 TIME-DIVISION MULTIPLEXING

A technique by which we can take advantage of the sampling principle for the purpose of time-division multiplexing is illustrated in the idealized representation of Fig. 14.3-1. At the transmitting end on the left, a number of band-limited signals are connected to the contact points of a rotary switch. We assume that the signals are similarly band-limited. For example, they may all be voice signals, limited to 3.3 kHz. As the rotary arm of the switch swings

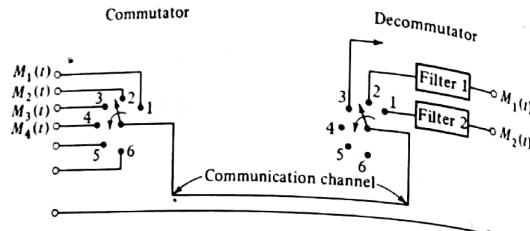


FIGURE 14.3-1
How the sampling principle can be used to transmit a number of band-limited signals over a single communications channel.

around, it samples each signal sequentially. The rotary switch at the receiving end is in synchronism with the switch at the sending end. The two switches make contact simultaneously at similarly numbered contacts. With each revolution of the switch, one sample is taken of each input signal and presented to the correspondingly numbered contact of the receiving-end switch. The train of samples at, say, terminal 1 in the receiver, passes through low-pass filter 1, and at the filter output the original signal $M_1(t)$ appears reconstructed. Of course, if f_m is the highest-frequency spectral component present in any of the input signals, the switches must make at least $2f_m$ revolutions per second.

When the signals to be multiplexed vary slowly with time, so that the sampling rate is correspondingly slow, mechanical switches, indicated in Fig. 14.3-1, can be used. When the switching speed required is outside the range of mechanical switches, electronic switching systems can be used. In either event, the switching mechanism, corresponding to the switch at the left in Fig. 14.3-1, which samples the signals, is called the *commutator*. The switching mechanism which performs the function of the switch at the right in Fig. 14.3-1 is called the *decommutator*. The commutator samples and combines the samples, while the decommutator separates samples belonging to individual signals so that these signals can be reconstructed.

The interlacing of the samples that allows multiplexing is shown in Fig. 14.3-2. Here, for simplicity, we have considered the multiplexing of just two signals, $M_1(t)$ and $M_2(t)$. Signal $M_1(t)$ is sampled regularly at intervals of T_s and at the times indicated in the figure. The sampling of $M_2(t)$ is similarly regular, but the samples are taken at a time different from the sampling time of $M_1(t)$. The input waveform to the filter numbered 1 in Fig. 14.3-1 is the train of samples of $M_1(t)$, and the input to the filter numbered 2 is the train of samples of $M_2(t)$. The timing in Fig. 14.3-2 has been deliberately drawn to suggest that there is room to multiplex more than two signals.

We observe that the train of pulses corresponding to the samples of each signal is modulated in amplitude in accordance with the signal itself. Accordingly,

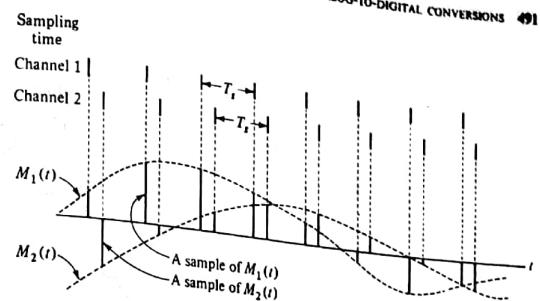


FIGURE 14.3-2
Interlacing two baseband signals.

the scheme of sampling is called *pulse-amplitude modulation (PAM)*. We note, however, that these pulses may vary continuously in amplitude. Hence PAM is an *analog* and not a *digital* system. We have discussed it here simply as a necessary prelude to *pulse-code modulation*, which is a digital system.

14.4 QUANTIZATION

The validity of the sampling theorem makes it possible to transmit or to process an analog signal by digital means. For we need not take note of the analog signal at all times but only at the sampling times, and hence in the intervals between samplings we shall have time to convert each sample voltage to digital form. The samples are continuously varying analog voltages. In digital form the allowable variation is not continuous since sample values must differ, at a minimum, by the least significant of the digits used in the digital representation. Hence the process of digitizing samples involves making an approximation. This process of approximation is called *quantization*.

The operation of quantization is illustrated in Fig. 14.4-1. A signal $M(t)$ is shown in Fig. 14.4-1a. This signal is the waveform V_i applied to the quantizer input. The output of the quantizer is called V_o . The quantizer has the essential feature that its input-output characteristic has the staircase form shown in Fig. 14.4-1b. As a consequence, the output V_o , shown in Fig. 14.4-1c, is the quantized waveform $M_q(t)$. It is observed that while the input $V_i = M(t)$ varies smoothly over its range, the quantized signal $V_o = M_q(t)$ holds at one or another of a number of fixed levels ... M_{-2} , M_{-1} , M_0 , M_1 , M_2 , Thus the signal $M_q(t)$ either does not change or changes abruptly by a quantum jump S called the *step size*.

The waveform $M'(t)$ shown dotted in Fig. 14.4-1c represents the output

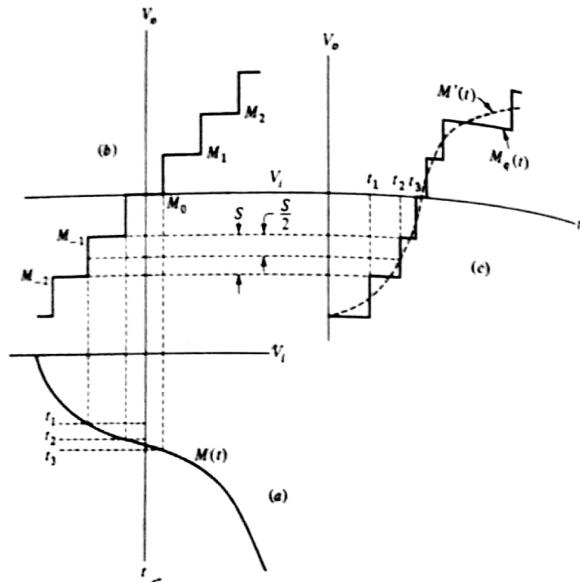


FIGURE 14.4-1

The operation of quantization. The step size is S . (a) The signal $M(t)$. (b) The input-output characteristic of the quantizer. (c) The quantizer output (solid line) response to $M(t)$. The dashed waveform $M'(t)$ shows the waveform of the output signal for a linear characteristic.

waveform, assuming that the output is linearly related to the input. If the factor of proportionality is unity, $V_o = V_i$ and $M'(t) = M(t)$. We see then that the level held by the waveform $M_q(t)$ is the level to which $M'(t)$ is closest. The transition between one level and the next occurs at the instant when $M'(t)$ crosses a point midway between two adjacent levels.

We see, therefore, that the quantized signal is an approximation to the original signal. The quality of the approximation can be improved by reducing the size of the steps, thereby increasing the number of allowable levels. Eventually, with small enough steps, the human ear or eye will not be able to distinguish the original from the quantized signal. To give the reader an idea of the number of quantization levels required in a practical system, we note that 512 levels can be used to obtain the quality of commercial color TV, while 64 levels gives only fairly acceptable color TV performance.

If we propose to quantize a signal with peak-to-peak range R and to use Q quantization levels, the step size S is determined by the condition that $QS = R$.

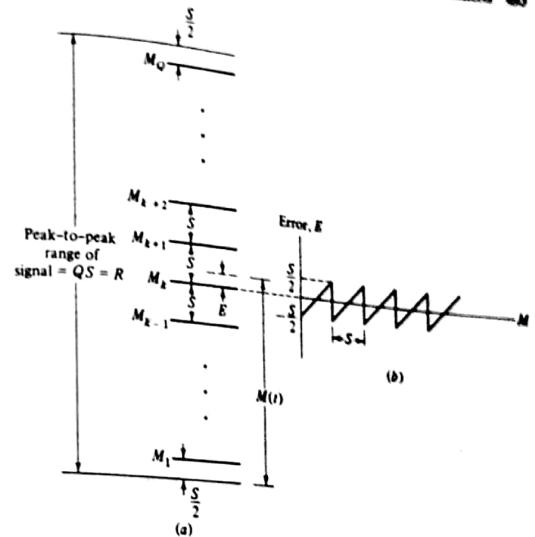


FIGURE 14.4-2

(a) A range of voltage over which a signal $M(t)$ makes excursions is divided into Q quantization ranges each of size S . The quantization levels are located at the center of the range. (b) The error voltage $E(t)$ as a function of the instantaneous value of the signal $M(t)$.

We would locate the quantization levels as indicated in Fig. 14.4-2a. In this way the maximum instantaneous quantization error would be $S/2$, as illustrated in Fig. 14.4-2b.

The complete process of digitizing an analog waveform is illustrated in Fig. 14.4-3. The signal $M(t)$ is regularly sampled at times indicated by the dots on the waveform. The anticipated peak-to-peak range R is 7 V extending from -3.5 to +3.5 V. We have allowed eight quantization levels located in such manner that maximum possible instantaneous quantization error is 0.5 V. Following common practice, we have assigned a set of binary digits to each level, using twos-complement representation (since there are eight levels, 3 bits are required).

The binary digits can now be transmitted or processed serially or in parallel. With serial processing and with three digits, as in the present case, the processing of each digit may occupy nominally one-third the interval between sampling times. With parallel processing, the entire interval is available for each bit.

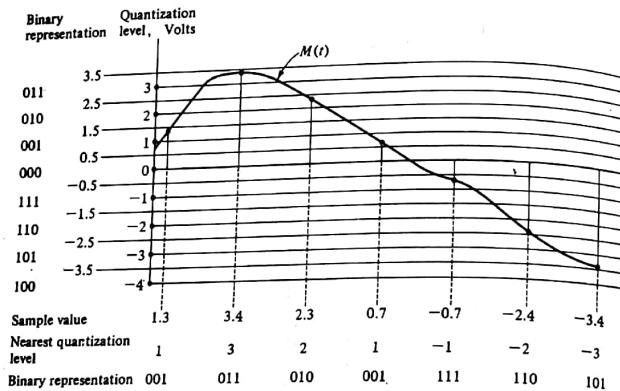


FIGURE 14.4-3

A message signal is regularly sampled. Quantization levels are indicated. For each sample the quantized value is given and its binary representation is indicated using two-complement representation.

Beginning with the next section we shall consider circuits for converting digital representations to analog form. The D/A converters are rather simpler than the A/D converters. Further, a D/A converter is frequently used within the structure of A/D converters. For these reasons we shall consider D/A converters first.

14.5 THE WEIGHTED-RESISTOR D/A CONVERTER

The structure of a weighted-resistor D/A converter is shown in Fig. 14.5-1. The input (not shown in the figure) is an N -digit binary signal $V = V_{N-1}V_{N-2}\cdots V_0$ in which each V_k is a voltage whose level is such as to represent either logic 1 or logic 0. It is assumed that all the V_k are available simultaneously, i.e., in parallel on N lines. If the V_k appear serially, the bits must be entered into a shift register so that all bits can be made available simultaneously. The voltages V_k normally available to drive the converter need not be precisely fixed voltages but only identifiable as representing one logic level or the other. For this reason the V_k 's are not applied directly to the converter but are used instead to operate the (electronic) switches S_0, S_1, \dots, S_{N-1} .

When V_k corresponds to logic 1 or 0, the switch S_k is thrown to the 1 or 0 position connecting a resistor of resistance R_k to the precisely controlled voltage

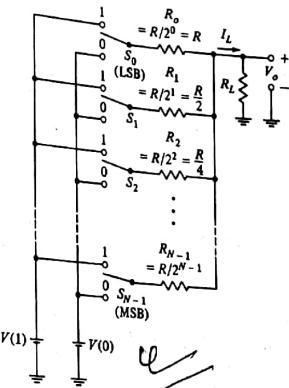


FIGURE 14.5-1
A weighted-resistor D/A converter.

source $V(1)$ or $V(0)$. The least significant bit (LSB) V_0 operates switch S_0 , which is connected to the output through the resistor of highest resistance $R_0 = R$. The most significant bit (MSB) operates S_{N-1} . Observe that the resistors R_0, R_1, \dots, R_{N-1} are weighted so that their resistances are inversely proportional to the numerical significance of the corresponding binary digit. The operation of the converter depends on the resistance values of successive resistors being related by a factor of 2 and does not depend on the absolute value of the resistors. Hence the parameter R in Fig. 14.5-1 is arbitrary.

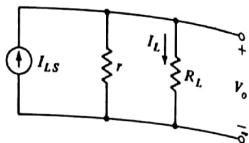
For simplicity, let us assume initially that $V(1) = V_R$, a fixed reference voltage, that $V(0) = 0$, that is, all 0 switch positions are grounded, and that the load $R_L = 0$ (in which case $V_o = 0$). Then the output current I_L is readily calculated in terms of the switch positions. Let $S_k = 1$ and $S_k = 0$ represent respectively the situation in which S_k is thrown to the 1 or 0 position. Then the load current I_{LS} (which, being the current into a short circuit, we call here I_{LS}) is

$$I_{LS} = V_R \left(\frac{S_{N-1}}{R_{N-1}} + \frac{S_{N-2}}{R_{N-2}} + \cdots + \frac{S_0}{R_0} \right) \quad (14.5-1a)$$

$$= \frac{V_R}{R} (S_{N-1} 2^{N-1} + S_{N-2} 2^{N-2} + \cdots + S_0 2^0) \quad (14.5-1b)$$

Thus we observe that I_{LS} has a numerical value which is directly proportional to the numerical value of the binary number $S = S_{N-1}S_{N-2}\cdots S_0$, the factor of proportionality being V_R/R .

FIGURE 14.5-2
The D/A converter resistor array of Fig. 14.5-1 replaced by a Norton's equivalent circuit.



This proportionality persists, albeit with a different factor of proportionality, if R_L is not equal to zero. That such is the case is to be seen from Fig. 14.5-2. Here we have replaced the converter network (excluding R_L) by its Norton's equivalent circuit. This equivalent circuit consists of a current source I_{LS} shunted by a resistor r , which is the output impedance of the network. The output impedance r is equal in resistance to the parallel combination of all the resistors R_0, R_1, \dots, R_{N-1} and, as can be verified (Prob. 14.5-1), is given by

$$r = \frac{R}{2^N - 1} \quad (14.5-2)$$

The new load current into R_L is

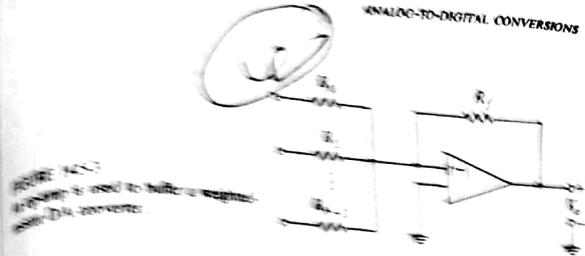
$$I_L = \frac{r}{r + R_L} I_{LS} \quad (14.5-3)$$

and the output voltage is, using Eqs. (14.5-1) to (14.5-3),

$$V_o = R_L I_L = \frac{R_L V_R}{R + (2^N - 1)R_L} (S_{N-1} 2^{N-1} + S_{N-2} 2^{N-2} + \dots + S_0 2^0) \quad (14.5-4)$$

With $V(1) = V_R$ and $V(0) = 0$, the range of output voltage V_o extends from 0 V when $S = 0 \dots 00$ to $V = R_L V_R / (r + R_L)$ when $S = 1 \dots 11$. However, by selecting $V(1)$ and $V(0)$ appropriately we can offset the range of V_o to suit our convenience. Thus, for example, with $V(1) = V_R$ and $V(0) = -V_R$, the excursion of V_o will extend in the negative as well as in the positive direction, being symmetrical about 0 V. Most commonly an op-amp is employed to buffer the D/A resistor network. Such an arrangement is indicated in Fig. 14.5-3. For simplicity the switches have been omitted from this diagram, but their presence at the input sides of R_0, R_1, \dots, R_{N-1} is to be understood. Because of the virtual ground at the inverting input terminal of the op-amp, the output is given by

$$V_o = -\frac{R_f V_R}{R_{N-1}} S_{N-1} - \frac{R_f V_R}{R_{N-2}} S_{N-2} - \dots - \frac{R_f V_R}{R_0} S_0 \quad (14.5-5)$$



PROBLEMS 14.5-1
A weighted-resistor array is indicated in Fig. 14.5-1 of the fact that $R_{N-1} = R/2^{N-1}$, etc.

$$V_o = -\frac{R_f R}{R} (S_{N-1} 2^{N-1} + S_{N-2} 2^{N-2} + \dots + S_0 2^0) \quad (14.5-6)$$

It is of some interest to note that when an op-amp is used, if the op-amp is ideal, it would not be necessary to use double-pole switches at the inputs to the converter. Single-pole switches would serve as well. Such a switch would connect the corresponding resistor to a voltage source V_i when the switch position was $S = 1$ and leave the resistor floating, i.e., disconnected at its midpoint when $S = 0$. Such an arrangement is allowable, it is known, because of the virtual ground at the op-amp input. With the virtual ground, the amplifier input current due to any switch being closed is independent of the settings of any of the other switches. However, as a matter of practicality, a floating input terminal is often a source of spurious disturbance, but, the system is unimportant, but, often unavoidable coupling, and this procedure is not usually followed.

A difficulty associated with the weighted-resistor converter stems from the wide range of resistor values which must be used. Suppose, for example, the resistor corresponding to the MSB is 1 kΩ and the converter is to accommodate 12 bits. Then the resistor associated with the least significant bit accommodates 11 bits. Then the resistor associated with the least significant bit is $1 \text{ k}\Omega \times 2^{11} = 1.2 \text{ M}\Omega$. Resistor accuracy < high- \times range which have the required precision and which total over a wide temperature range are difficult to produce. It is especially difficult in a monolithic fabrication, and is one of the difficulties of interest in integrated-circuit fabrication. An approach to the problem is an alternative converter-resistor array which circumvents this difficulty.

14.6 THE R - 2R LADDER DA CONVERTER

The converter-resistor array of Fig. 14.5-2 has resistors of only two sizes, R and $2R$. In this figure only 4 bits are indicated. As a consequence the array can be expanded to accommodate an arbitrary number of bits. Note that this requires twice as many resistors for the same number of input bits as the weighted-resistor array. Again, for simplicity in Fig. 14.6-1 the details of the

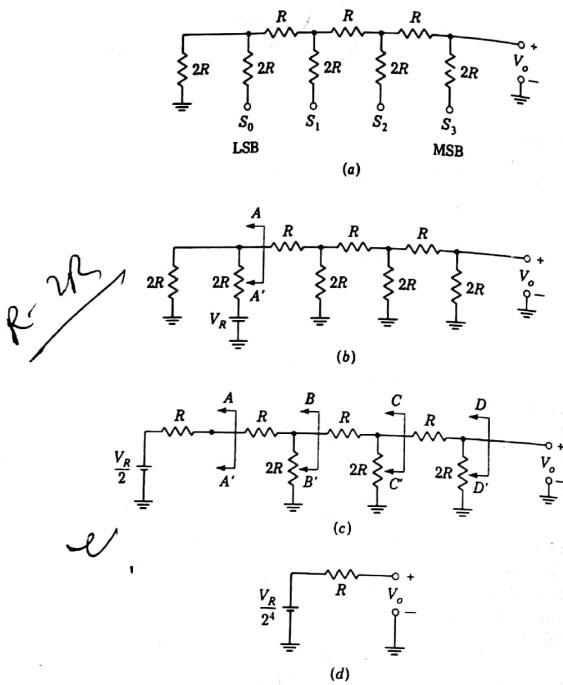


FIGURE 14.6-1
(a) The R-2R ladder D/A converter. (b and c) Thevenin's theorem used to determine the output voltage V_o . (d) The final equivalent circuit as seen at the output.

switches have been omitted, but it is to be understood that when $S_k = 1$, the corresponding resistor is connected to a voltage V_R and when $S_k = 0$, the resistor input is grounded.

To see most simply the relative weight at the output of the individual switches, consider the situation in Fig. 14.6-1b, where we have set $S_0 = 1$ while $S_1 = S_2 = S_3 = 0$. In this example we find the voltage source V_R in series with a resistor $2R$. Applying Thevenin's theorem at AA' , we find, as shown in Fig. 14.6-1c, that we now have a voltage source $V_R/2$ in series with R . We repeat the application of Thevenin's theorem at BB' , CC' , and DD' . We find that at each

such application the voltage source is again divided by 2, while the Thevenin's equivalent output impedance remains constant at R . The final equivalent circuit for the output is shown in Fig. 14.6-1d. If we repeated these operations starting with $S_1 = 1$, $S_3 = S_2 = S_0 = 0$, we would find an equivalent circuit as in Fig. 14.6-1d except that the voltage source would be $V_R/2^3$, and so on for switches S_2 and S_3 . Thus, at the output, each switch contributes its proper relative binary weight. For the arrangement in Fig. 14.6-1a we then have

$$V_o = V_R \left(\frac{S_3}{2^1} + \frac{S_2}{2^2} + \frac{S_1}{2^3} + \frac{S_0}{2^4} \right) \quad (14.6-1a)$$

$$V_o = \frac{V_R}{2^4} (S_3 2^3 + S_2 2^2 + S_1 2^1 + S_0 2^0) \quad (14.6-1b)$$

or, more generally, if there are N input digits and correspondingly N switches, we have

$$V_o = \frac{V_R}{2^N} (S_{N-1} 2^{N-1} + S_{N-2} 2^{N-2} + \dots + S_0 2^0) \quad (14.6-2)$$

Equation (14.6-2) does not take into account a possible load placed across the output. Such a load will change the absolute value of the output, but, as with the weighted-resistor converter, it will not change the relative weight of the individual switches. For suppose that the load is R_L . Then, referring to Fig. 14.6-1d, the output when $S_0 = 1$ will be

$$V_o = \frac{V_R}{2^4} \frac{R_L}{R + R_L} \quad (14.6-3)$$

However, no matter what the position of the switches, the Thevenin's output impedance of the converter is R , and hence the effect on a load R_L is in every case simply to reduce the output by the factor $R_L/(R_L + R)$.

Usually the R-2R ladder converter array is constructed as shown in Fig. 14.6-2a. Note here the inclusion of an additional $2R$ load resistor to the right of the switch corresponding to the most significant digit. The feature (sometimes an advantage) which results from the inclusion of this extra resistor is that the impedances seen looking into each input of the array are the same. As indicated, this impedance is $3R$. A D/A converter using an R-2R ladder with the extra resistor and using an op-amp buffer is shown in Fig. 14.6-2b. It can be verified that the output V_o (except for sign) is given by Eq. (14.6-2).

The weighted-resistor converter assigns weights to digital input bits by using appropriately weighted resistors. In the R-2R ladder, bits are weighted by providing paths for current division with consequent successive attenuations for bits of lower significance. A compromise between these two schemes is possible. One such compromise circuit is shown in Fig. 14.6-3. As usual, S_0 is the least significant bit and S_7 the most significant bit. We have chosen to arrange the bits in groups of four. Within each group the bits are introduced through weighted resistors whose resistances are related by powers of 2, as required. It

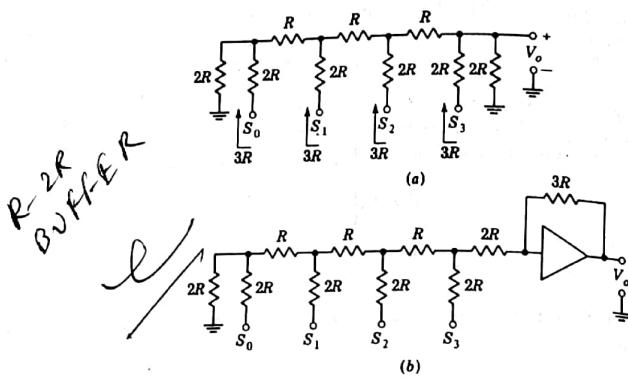


FIGURE 14.6-2
Connection of the R-2R ladder to an op-amp.

is, however, required that there be an additional attenuation of 16 between the group $S_0 S_1 S_2 S_3$ and the group $S_4 S_5 S_6 S_7$. Thus when $S_3 = 1$, the consequent input current to the virtual ground of the op-amp should be one-sixteenth the current which results when $S_7 = 1$. Similarly this same relationship should hold between S_2 and S_6 , etc. The resistor r has been inserted between the two groups to provide this attenuation. It can be verified that $r = 8R$ will reduce the current by the factor $\frac{1}{16}$, as required for a straight binary input format.

The arrangement of Fig. 14.6-3 is also very convenient to accommodate a BCD input format. In this format, the four bits $S_3 S_2 S_1 S_0$ represent a decimal digit 0 to 9 in the units position while $S_7 S_6 S_5 S_4$ represent a decimal digit in the tens position. In such cases the resistor r must be selected to provide an attenuation by 10 rather than by 16. It can be verified (Prob. 14.6-3) that such attenuation by 10 requires that $r = 4.8R$.

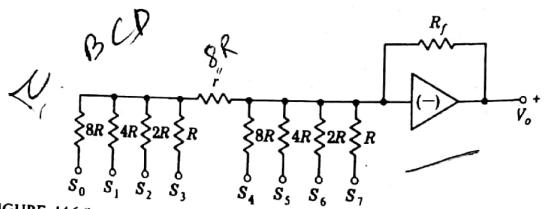


FIGURE 14.6-3
A D/A converter which effects a compromise between the use of weighted resistors and the use of a ladder network.

14.7 SWITCHES FOR D/A CONVERTERS

In Chap. 1 we discussed the use of diodes and transistors (both bipolar and FET) as switches. In Chap. 13 we discussed the use of these devices in analog-signal switching circuits. These devices can be used as well to serve as the switches in D/A converters.

Turned-ON FETs behave like simple resistors. When such devices are used, either the resistors in the converter array must be large enough for the account. In the latter case it may be necessary to provide some measure of temperature compensation (see Prob. 14.7-1) since the FET resistance is temperature-sensitive. In saturation, bipolar transistors have negligible resistance. Since the offset voltage is appreciably smaller when the transistor is used in the inverse direction (see Fig. 1.10-1), this inverse operation would normally be used.

There is, however, a difficulty associated with using bipolar transistor switches in the inverse direction in a converter circuit. To appreciate the difficulty, consider the situation represented in Fig. 14.7-1. Here the two transistors are to serve as the double-pole switch in, say, the converter circuit of Fig. 14.5-1, in which we have set $V(1) = V_R = 5V$ and $V(0) = -V_R = -5V$. The resistor R_e represents one of the converter resistors. The side of R_e distant from the switches is considered to be at ground by virtue of its connection to the virtual ground at the input of an op-amp. Each transistor is being used in the inverse direction with the normal collector junction serving here as the emitter junction. We shall cut off a transistor by arranging that the voltage across its base-collector junction shall be 0 V, and we shall saturate the transistor by forward biasing its base-collector junction to the extent of 0.75 V. Then the voltages indicated on the figure for V_{B1} and V_{B2} are appropriate to saturate one transistor while the other is OFF and vice versa. In this way the ungrounded side of R_e is connected to -5 or $+5$ V respectively.

Now consider, say, that $V_{B1} = -5$ V and $V_{B2} = 4.25$ V, so that $T1$ is OFF and $T2$ is ON. Then the voltage between the base and the emitter of $T1$ is very

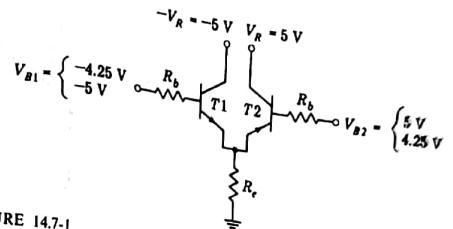


FIGURE 14.7-1
Bipolar transistors used in the inverse manner as switches in a D/A converter.

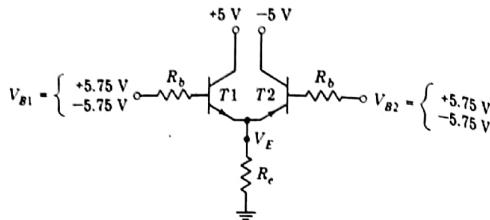


FIGURE 14.7-2
Overdriven emitter followers used as switches in a D/A converter.

nearly $V_{BE1} = -9.25$ V. This large a voltage between base and emitter of a transistor is larger than a transistor can normally sustain. For this reason the switching arrangement indicated in Fig. 14.7-1 is not acceptable.

An alternative switching arrangement which circumvents this voltage difficulty is indicated in Fig. 14.7-2. Here the transistors are used in the normal rather than in the inverted fashion, but the circuit configurations are those of emitter followers. As pointed out in Sec. 1.10, an overdriven emitter follower can have an offset voltage of 0 V since the saturation factor σ is negative. In Fig. 14.7-2 the bases are driven between about +5.75 and -5.75 V. In the first case $T1$ is in saturation and $V_E \approx 5$ V while $T2$ is off. The base-emitter voltages are the same with $V_{BE1} = V_{BE2} = 0.75$ V. In the second case $T2$ is on, and $T1$ is off. The emitter voltage V_E is now $V_E \approx -5$ V. We note as a possible disadvantage of the present arrangement that large base-voltage swings (11.5 V) are required.

An example of a D/A converter As an example of a D/A converter using the switching scheme of Fig. 14.7-2 we consider the circuit of Fig. 14.7-3. Although a weighted-resistor converter is indicated, the scheme is suitable for a ladder converter as well. The reference voltages $\pm V_R$ are ± 5 V. Thus, neglecting the saturation voltage across $T4$ or $T3$, we have switch voltages of ± 5 V. To see that the collector-emitter voltage of $T4$ (and of $T3$) is indeed negligible assume that $T2$ is cut off. Then $T4$ is on. The emitter voltage of $T4$ is therefore approximately 5 V, and the base voltage V_{B4} is approximately 5.75 V. Thus, current $I_{E4} \approx (12 - 5.75)/(4.4 \text{ k}\Omega) \approx 1.4$ mA. However, the emitter current $I_{E4} \approx 5/(10 \text{ k}\Omega) = 0.5$ mA. Note therefore that $I_{C4} \approx -0.9$ mA and $\sigma \equiv I_C/h_{FE} I_B \approx -0.6/h_{FE}$. It may be shown from Eq. (1.10-3) that when $\sigma \approx -0.6/h_{FE}$ that $V_{CE} \approx V_T \ln(1 + 0.4/h_{FE})$. If we assume that $h_{FC} = 0.1$ we then find that $V_{CE} \approx 40$ mV.

The input driving voltage swings from about 0 V to about 3 V. With the input at 3 V, $T1$ and $T2$ are off, and $T4$ is driven to saturation by base current derived from the +12-V source and flowing through the 2.2-k Ω

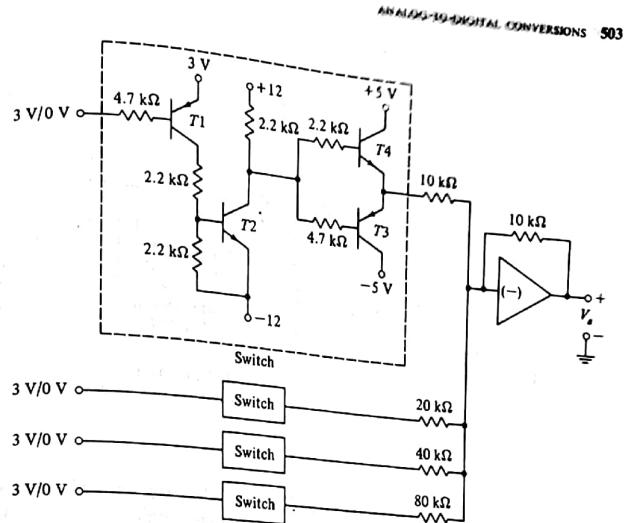


FIGURE 14.7-3
An example of a D/A converter.

collector resistor of $T2$ and the 2.2-k Ω base resistor of $T4$. While $T4$ is on, $T3$ is off. When the input is at 0 V, both $T1$ and $T2$ saturate, as can be verified (Prob. 14.7-1). In this case, $T3$ is driven to saturation by base current ($T2$) and through the 4.7-k Ω base resistor of $T3$. In this case, assuming that $V_{CE2} \approx 0.2$ V, we find $I_{B3} \approx 1.3$ mA and, hence, σ remains $\sigma \approx -0.6/h_{FE}$. Now the emitter voltages of $T3$ and $T4$ are again very nearly at -5 V.

14.8 A CURRENT-DRIVEN D/A CONVERTER

A D/A converter using the ladder array of resistors and using current sources is shown in Fig. 14.8-1. We assume that the switches are of the make-before-break type so that the current I need never be interrupted. The present network is equivalent to the former network of Fig. 14.6-2. In the former network a voltage source is applied, as required, in series with resistors of resistance $2R$. In the present network the series combinations of voltages and resistors are replaced by the combination of a current source in parallel with resistors of resistance $2R$.

The current-driven converter has a potential merit in comparison with the voltage-driven converter. The voltage-driven converter requires transistor

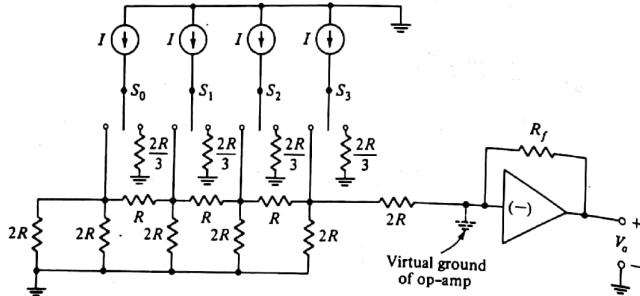


FIGURE 14.8-1
A current-driven ladder D/A converter.

switches which are driven to saturation. The time delay associated with bringing a transistor out of saturation often establishes an upper limit to the speed with which such voltage-driven converters can operate. On the other hand, in current-driven converters it is feasible to use a switching arrangement in which transistors are not driven into and out of saturation; instead, we can use an arrangement in which a current is switched into and out of the converter-resistor array, devising thereby a technique to drive transistors from the active region to cutoff and vice versa. We have already encountered such a switching scheme in the basic difference-amplifier configuration of emitter-coupled logic.

A current-driven ladder converter using the difference-amplifier switches encountered in emitter-coupled logic is shown in Fig. 14.8-2. Currents are switched into the ladder when the transistors T_{0A} , T_{1A} , T_{2A} , or T_{3A} conduct, and the currents are diverted to transistors T_{0B} , T_{1B} , T_{2B} , or T_{3B} when they conduct. The circuit shown in Fig. 14.8-2b provides the base-biasing voltage V_B for all the switches. The converter circuit is intended to be used in conjunction with ECL gates. At room temperatures these gates (see Sec. 7.4) have logic levels -0.76 V (logic 1) and -1.58 V (logic 0). Consequently, the base voltage V_B is set very nearly midway between these levels at $V_B = -1.15$ V. Thus, when, say, $V_0 = -1.58$ V, T_{0B} is OFF and T_{0A} is ON, and when $V_0 = -0.76$ V, T_{0B} is ON and T_{0A} is OFF.

We assume that at room temperature the voltage across diodes $D1$ and $D2$ as well as across the base-emitter junction of a conducting transistor is 0.75. On this basis, it can readily be verified that because the supply voltage $-V_{EE}$ in the biasing circuit has been selected at $-V_{EE} = -8.1$ V, the base bias V_B turns out to be $V_B = -1.15$ V. For we have

$$V_B = -V_{EE} + V_Z + V_{D2} + V_{D1} - V_{BE}(TB) \quad (14.8-1)$$

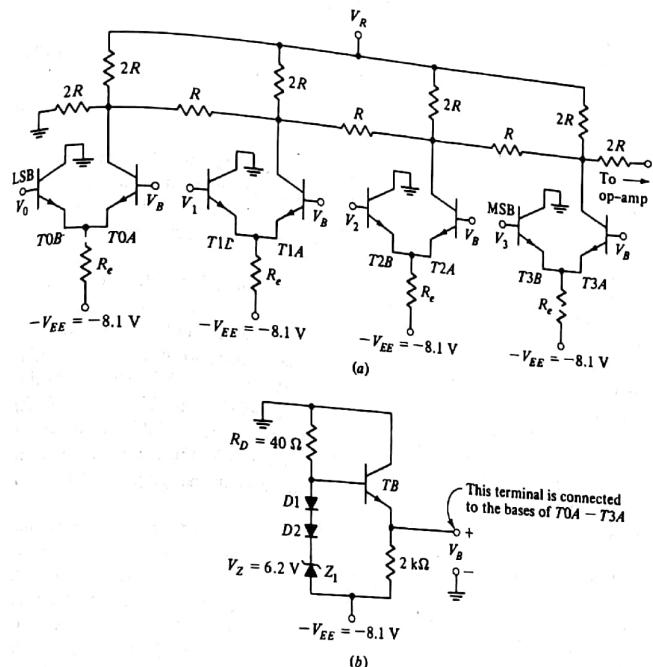


FIGURE 14.8-2
(a) A high-speed current-driven D/A converter. (b) Base-biasing circuit.

where V_Z = voltage across zener diode
 V_{D1} , V_{D2} = voltages across diodes $D1$ and $D2$
 $V_{BE}(TB)$ = base-to-emitter voltage of TB
We then find

$$V_B = -8.1 + 6.2 + 0.75 + 0.75 - 0.75 = -1.15 \text{ V} \quad (14.8-2)$$

The circuitry also provides a measure of temperature compensation in that the current injected into the ladder at each switch is independent of the temperature. In each case this injected current is the collector current of the corresponding transistor. This collector current, in turn, except for a small and constant fraction due to base current, is equal to the emitter current. The voltage across the emitter resistor of, say, $T0$ is

$$V_{R_e} = V_B - V_{BE}(T0A) + V_{EE} \quad (14.8-3)$$

Using Eq. (14.8-1) and assuming $V_{D1} = V_{D2} = V_{BE}(TB) = V_{BE}(T0A)$, we find V_{R_e} and the current I_{R_e} to be

$$V_{R_e} = V_Z \quad (14.8-4)$$

and

$$I_{R_e} = \frac{V_Z}{R_e} \quad (14.8-5)$$

In Sec. 1.4 it was pointed out that depending on the operating voltage of a zener diode, an operating current can generally be found at which the zener-diode voltage is temperature-independent. The zener diode in the bias circuit is so selected, and hence Eq. (14.8-5) indicates that the injected current is relatively independent of temperature.

We may consider now various details in connection with the circuit of Fig. 14.8-2, all of which combine to allow the circuit to operate at high speed. We have already noted that the transistor switches do not go to saturation and that storage-time delays are thereby avoided. Next we note that assuming the base-emitter voltage of a conducting transistor to be 0.75 V, we can readily verify that when either of the transistors in the switch pair ($T0A$ or $T0B$, for example) is off, it operates with a base-emitter voltage of 0.35 V. Since we have always assumed that cut-in occurs at about 0.65 V, the transistor is never deeply into cutoff and there need be no long delay in carrying the transistor from cutoff into the active region.

When a switch operates to inject current into a node of the ladder, the voltage at that node (and to a lesser extent at all other nodes) must change. There are the inevitable stray shunt capacitances present in the ladder network. These capacitances must charge and discharge, in response to the operation of the switches, with a consequent slowing of the operation of the converter. To minimize this speed limitation these stray capacitances must find themselves in circuits having a small time constant. For this reason the resistors of the ladder must be small. Resistors $R = 50 \Omega$ and $2R = 100 \Omega$ are not unusual.

There is, however, a problem associated with the circuit of Fig. 14.8-2a: the current injected by $T0A$ takes longer to reach the op-amp than the current injected by, say, $T3A$ since the ladder network appears to be a lossy transmission line at very high frequency. This time difference can result in a rather large spike in the op-amp output. To see how this can arise consider that the digital input is 1000 (8 V) at $t = 0$ and somewhat later switches to 0111 (7 V). We further assume that the switching occurs at the same time throughout the circuit. However, due to the propagation delay in the ladder network, we find that $T3A$ switches first, so that the 8-V level immediately drops to 0 V. The voltage then rises to 4, 6, and finally 7 V.

To reduce the duration of the spike, a very-high-speed D/A converter can arrange to apply the logic switch voltages to $T0B$ before $T1B$, etc. In a later section we discuss the need to offset the output of a D/A to encompass a range of voltages which are both positive and negative. The reference voltage V_R in Fig. 14.8-2a is used to provide for such an offset.

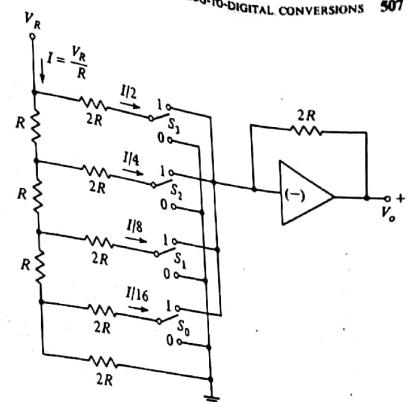


FIGURE 14.9-1
The inverted-ladder D/A converter.

14.9 THE INVERTED-LADDER D/A CONVERTER

We noted above that the D/A converter shown in Fig. 14.8-2 produces a spike during conversion as a result of the propagation delay suffered in the ladder. In the inverted ladder shown in Fig. 14.9-1 the switches are connected directly to the op-amp, thereby eliminating the propagation-delay-time problem.

In the inverted ladder the switches are located at the input to the op-amp rather than at the reference source V_R , as in Fig. 14.8-2. A 4-bit converter is indicated, but the extension to an arbitrary number of bits is obvious. Keeping in mind that at the input to the op-amp we see a virtual ground, it is clear that the currents that flow in the ladder are independent of the switch position. For in either position of the switch, the $2R$ resistors are connected to ground. We assume that the switches are of the make-before-break type, so that at no time is the switch arm disconnected from ground or virtual ground. It can now be readily verified that the current drawn from V_R is $I = V_R/R$ and that the currents in the individual $2R$ resistors are related by powers of 2, as shown. The switches serve to direct these currents into the op-amp or to direct the currents to ground. It can also be verified that the output of the amplifier V_o is given by

$$V_o = \frac{V_R}{2^3} (S_3 2^3 + S_2 2^2 + S_1 2^1 + S_0 2^0) \quad (14.9-1)$$

The important merit of the inverted-ladder configuration is that the currents through the resistors of the ladder do not change with switching. Hence the

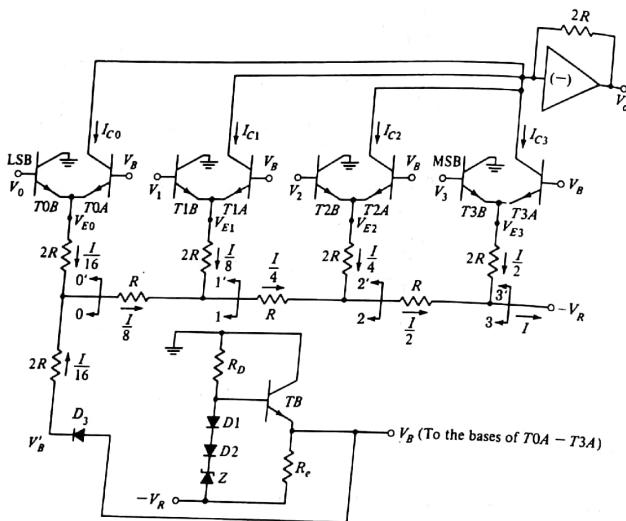


FIGURE 14.9-2
A high-speed D/A converter using the inverted-ladder configuration.

voltages across these resistors do not change, and there is no delay occasioned by the necessity for parasitic capacitors to charge or discharge. There is a second advantage. In the normal ladder the switches must switch the full reference voltage in and out of the circuit. In the inverted ladder the switches (thanks to their location) never have to sustain any appreciable voltage.

An example of an inverted-ladder converter is shown in Fig. 14.9-2. Here again the ECL-type difference amplifier has been used as a switch in order to avoid the delay associated with bringing a transistor out of saturation. We may note the correspondences and also the small differences between the circuits of Figs. 14.9-2 and 14.9-1. In Fig. 14.9-2 the reference voltage is $-V_R$ rather than V_R simply because npn transistors have been used. More important, we note that the last resistor in the ladder, which in Fig. 14.9-1 is connected to ground, is connected in the present circuit to the voltage V'_B . This connection, as we shall see, serves two functions. On the one hand it connects this last resistor to a voltage equal to the voltage V_{E0} ($= V_{E1} = V_{E2} = V_{E3}$) to which the other $2R$ resistors are connected. In Fig. 14.9-1 these resistors are all returned to the

same point, namely, ground. The ground connection is required there since the switches are assumed ideal and this common point must be at the same potential as the virtual ground at the op-amp input. In Fig. 14.9-2 the switches are not perfect, and the common voltage is not the voltage at the op-amp input. The second function this connection provides is a measure of temperature compensation.

Let us consider that in each case in Fig. 14.9-2 the A transistor is ON. Then the B transistor is OFF, and the voltages at the common emitters are

$$V_{E0} (= V_{E1} = V_{E2} = V_{E3}) = V_B - V_{BE} \quad (14.9-2)$$

in which V_{BE} is the voltage drop from base to emitter of a transistor (T0A, T1A, T2A, or T3A). We also find, referring to TB, that

$$V_A = V_B - V_{D3} \quad (14.9-3)$$

If we now assume that the voltage V_{B3} across the diode is the same as V_{BE} , then $V_{E0} = V'_B$, as required. On this basis the currents through the two leftmost resistors $2R$ in the circuit are the same, and we have assigned to them the value $I/16$. Following now the pattern of Fig. 14.9-1, i.e., noting that the circuits seen looking into 0-0', 1-1', 2-2', 3-3' each consist of an open-circuit voltage source V'_B in series with a resistor R , it is apparent that the currents assigned to the other resistors of the ladder are correct.

Now we can also verify that the currents in the ladder, and hence the currents injected into the op-amp, are independent of the temperature. These currents depend on the common voltage $V_B = V_{E0} = V_{E1} = V_{E2} = V_{E3}$. Calculating V'_B , (see transistor TB) we find

$$V'_B = -V_R + V_Z + V_{D1} + V_{D2} - V_{BE}(TB) - V_{D3} \quad (14.9-4)$$

Again assuming that the diode voltages are equal and equal also to V_{BE} , we find

$$V'_B = -V_R + V_Z \quad (14.9-5)$$

which, as discussed in connection with the circuit of Fig. 14.8-2, can be adjusted to be temperature-independent.

There is a source of error in the circuit of Fig. 14.9-2. To explore this point assume initially that all the A transistors are ON. Now let V_0 increase in voltage in order to transfer current from T0A to T0B. If T0A is to turn OFF, the emitter voltage V_{E0} must rise. As a consequence, the current in the $2R$ emitter resistor of T0 must change, and thereby all the other currents (except in the emitter resistor of T3) will be disturbed. Since these currents are required to remain constant, an error will result. However, in a practical case the error need not be large enough to be serious. The calculation of this error in a typical case is left as an exercise (see Prob. 14.9-2).

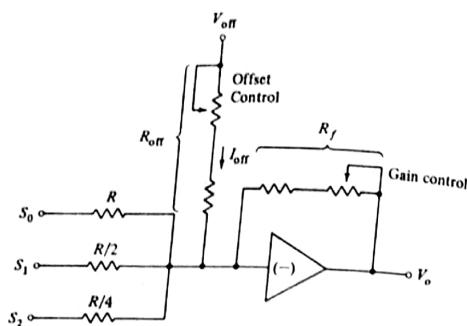


FIGURE 14.10-1
Circuit used to offset the output voltage of a D/A converter.

14.10 INPUT AND OUTPUT FORMATS OF A D/A CONVERTER

In the circuit arrangement of Fig. 14.5-1 suppose that we set $V(0) = 0$ V and deliver the current I_L to an operational amplifier, as in Fig. 14.5-3. Then when $S = S_{N-1} \cdots S_0 = 0 \cdots 0$, that is, all zeros, V_o will be $V_o = 0$, and when $S = 1 \cdots 1$, that is, all ones, V_o will be given by Eq. (14.5-6). If we take $V_R = 1$ V and $R = R_f$ and assume, say, a 3-bit converter, the output range will extend from 0 to -7 V. We describe the output as *unipolar* since it swings in only one direction. Since the input bits have a straightforward binary significance (each bit has the numerical significance of some power of 2), the overall D/A converter is described as operating in the *binary unipolar format*. If, however, $V(0) \neq 0$, we can *offset* the output swing. For example, if $V(1) = \frac{1}{2}$ V and $V(0) = -\frac{1}{2}$ V, the output swing extends symmetrically about 0 V from -3.5 to +3.5 V. In this case the format is called *bipolar binary* or *offset binary*.

An alternative method of offsetting the output range of a D/A converter is shown in Fig. 14.10-1. Here we have made provision to inject an offset current I_{off} into the input of the amplifier. Also included in the circuit is an adjustment of the gain of the amplifier. It is to be noted that because of the virtual ground at the amplifier input, the adjustment of the gain control and the offset control are independent of each other. Further, and again because of the virtual ground, the addition of the offset control has no effect on the input current of the amplifier which is due to the resistor array of the converter. The offset produced in the output voltage V_o is $V_o = -(R_f/R_{off})V_{off}$.

As long as we do not intend to use the representation to perform arithmetic operations, the method of offsetting provides a useful and convenient representation of negative numbers. For example, suppose, as above, that we start with a

3-bit converter which yields $V_o = 0$ V for an input 000 and $V_o = 7$ V for an input 111 [use $V(1) = -1$ V and $V(0) = 0$]. Now, suppose that we adjust the offset voltage so that $V_o = 0$ V for an input 100. Then the digital input and analog output will be related as in Table 14.10-1. Here, 000 represents -4; 001 represents -3, and so on. Such a representation of negative numbers would be most convenient and acceptable if, say, the analog output were intended to deflect the pointer of a meter one way or another, depending on the sign of the input number to the D/A converter. Note that for N bits, 2^N is an even number. Hence the offset cannot be adjusted so that the output range is exactly symmetrical about 0 V if it is required that one digital input correspond to the analog output 0 V.

Suppose, on the other hand, the digital representation of negative numbers we must deal with is a representation more suitable for arithmetic computation, say the two's-complement representation. A procedure which will allow a D/A converter to read and convert such a number representation is given in Table 14.10-2.

As can be seen from Table 14.10-2 and taking Table 14.10-1 into account, to convert from a two's-complement format to an analog signal we apply to the

Table 14.10-1

Digital input	Analog output V_o
111	+3
110	+2
101	+1
100	0
011	-1
010	-2
001	-3
000	-4

Table 14.10-2

Decimal	Two's-complement representation	Offset binary format when 100 is adjusted for 0 V
+3	011	111
+2	010	110
+1	001	101
0	000	100
-1	111	011
-2	110	010
-3	101	001
-4	100	000

D/A converter the digital input with the most significant bit complemented and then adjust the offset so that 100 corresponds to 0 V.

If the digital input to the converter is presented in a ones-complement format, the appropriate conversion procedure is apparent from an examination of Table 14.10-3.

Observe that the ones-complement representations of the positive numbers +0 to +3 are identical to the representations of the unipolar binary numbers. Therefore the digital inputs representing positive numbers are applied directly to the D/A converter, which is set for a zero offset voltage.

If the ones-complement representation of the negative numbers 111 to 100 were applied directly to the converter, the output analog voltage would read +7 to 4 V. To shift these voltages we apply a -7-V offset voltage to the op-amp whenever the sign bit in the ones-complement format is 1. For example, the number 110 (in ones complement) when applied to a D/A converter would read +6 V rather than -1 V. By applying a -7-V offset the correct analog voltage $+6 - 7 = -1$ V is read.

One way we can conveniently arrange for the correct offset is indicated in Fig. 14.10-2. Here we use symmetrical reference voltages $V(1) = -\frac{1}{2}$ V and $V(0) = +\frac{1}{2}$ V. We also use $V(1)$ and $V(0)$ as two offset voltages (V_{off} in Fig. 14.10-1) and add an additional switch, S_{off} . The offset resistor R_{off} is set to be equal to the parallel combination of R , $R/2$ and $R/4$, and finally we arrange that $S_{off} = \bar{S}_2$, so that when S_2 , the switch corresponding to the most significant digit (the sign bit), is at logic 1, S_{off} is at logic 0, and vice versa. Because of the symmetrical input reference voltages V_o would range between $\frac{3}{2}$ and $-\frac{3}{2}$ V if the additional offset provided were ignored. Hence we start with an offset of $-\frac{3}{2}$ V. When $S_2 = 0$, $S_{off} = 1$ and an additional offset of $+\frac{3}{2}$ is

Table 14.10-3

Decimal	Ones-complement representation	Unipolar binary
+7		111
+6		110
+5		101
+4		100
+3	011	011
+2	010	010
+1	001	001
+0	000	000
-0	111	
-1	110	
-2	101	
-3	100	

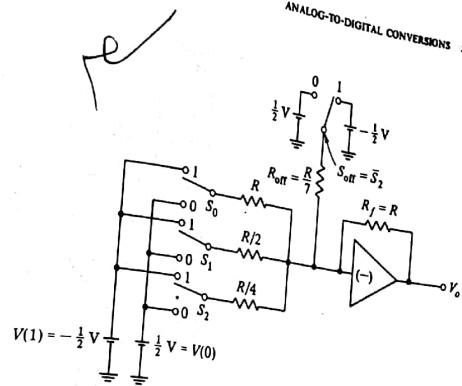


FIGURE 14.10-2
A D/A converter which accepts digital inputs in the ones-complement representation.

introduced, making the total offset zero. When $\bar{S}_2 = 1$, $S_{off} = \bar{S}_2 = 0$ and the offset voltages combine, making the total offset -7 V, as required.

It is left as a problem to show that the switch $S_{off} = \bar{S}_2$ can be eliminated by replacing S_2 by \bar{S}_2 and its resistance $R/4$ by $R/3$.

14.11 SPECIFICATIONS FOR D/A CONVERTERS

We consider now a number of the parameters which serve to describe the quality of performance of a D/A converter. These parameters are generally specified by manufacturers of converters.

Resolution This term specifies the number of bits the converter can accommodate and correspondingly the number of output voltages (or currents). For example, a converter which can accept 10 input bits is referred to as a converter with a 10-bit resolution. The number of possible output voltages is $2^{10} = 1,024$. Hence the smallest possible change in output voltage is $\frac{1}{1024}$ of the full-scale output range. Approximating 1,024 as 1,000, we can describe the resolution as being 1 part in 1,000, or 0.1 percent.

Linearity In an ideal D/A converter equal increments in the numerical significance of the digital-input should yield equal increments in the analog output. The *linearity* of a converter serves as a measure of the precision with which this requirement is satisfied. Linearity is measured in the manner suggested by Fig. 14.11-1. Assuming a unipolar binary input format, we have located the

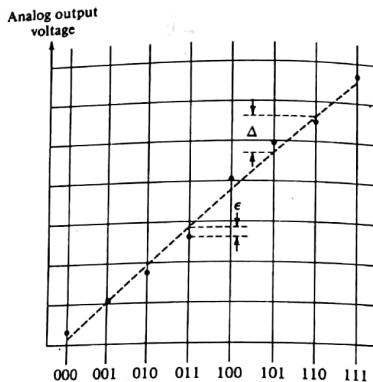


FIGURE 14.11-1
The measurement of linearity.

input bit combinations with fixed-interval separations, in the order of numerical significance along the horizontal axis. Along the vertical axis we have indicated by dots the corresponding analog output voltage for each case, as might be encountered in a physical converter. If the converter were perfect, the dots would fall on a straight line. In Fig. 14.11-1 we have drawn a straight line best fit to the dots and in a typical case indicated by ϵ the linearity error. The voltage Δ is the nominal analog output change corresponding to a digital input change equivalent to a change in the least significant bit (LSB).

The linearity of a converter is generally specified by comparing ϵ to Δ . Thus, commonly, we find the linearity of a commercial unit specified as "less than $\pm \frac{1}{2}$ LSB," meaning that $|\epsilon| < \frac{1}{2}\Delta$. This is a very important specification. For suppose that at one digital input we find that ϵ is positive and $\epsilon > \frac{1}{2}\Delta$ while at the next higher digital input ϵ is negative and $|\epsilon| > \frac{1}{2}\Delta$. In this case the converter would have the unacceptable feature of not being *monotonic*; i.e., an increase in digital input would yield a decrease in analog output.

The linearity of a converter depends principally on the accuracy of the resistors. It depends as well on the precision with which the voltage drops across the switches are fixed. Since both resistors and switch voltages are temperature-dependent, linearity may be adversely affected by substantial temperature changes.

Accuracy The *accuracy* of a converter is a measure of the difference between the actual analog output voltage and what the output should be in the ideal case. Lack of linearity contributes to inaccuracy. Further limitations on

FIGURE 14.11-2
Settling time and noise.

accuracy are contributed by the uncertainty in the reference voltages, the op-amp. amplifier noise, etc. A typical maximum specification for a low-quality converter might read "12-bit resolution, $\pm 1\text{ LSB}$ ".

Settling time When the digital input to a converter changes, switch capacitance and diode voltage changes appear. Because of the non-zero load resistance and inductance present in the passive circuitry, the transitions do not occur instantaneously. At an appreciable time additional transient voltage noise due to the characteristics of the active devices associated transients occurs. Therefore, after the change in digital input voltage as a function of time might be as shown in Fig. 14.11-2. Note that not only is there a time delay required to reach the new output level, but also an oscillation may occur. The time taken to settle from the input change to the time where the output no longer shows any change in its final value is called the *settling time*. The settling time depends strongly on the things on how we define "close enough." Typically, a good D/A converter might have a settling time greater than 100 ns to 1 μ s.

Some of the largest transients (often, unfortunately, greater than the settling duration) are produced by the operation of the switches. For example, suppose in a 4-bit unipolar converter the switches S_1, S_2, S_3, S_4 correspond to which the output is $+5\text{ V}$. Suppose for some change requires that now S_3, S_2, S_1, S_4 become **ON**, corresponding to 8 V . If we suppose that S_3 changes before the other switches have done so, that is, a best interval, we shall have $S_3, S_2, S_1, S_4 = 1111 = 15\text{ V}$. If the spikes produced in the D/A converter are objectionable, the D/A output voltage can be sampled and held. This new output is then low-pass filtered. It is found in practice that the glitches formed by the S/H circuit contain significantly less energy than the typical D/A spike. Thus, the noise produced after filtering the S/H circuit is much less than the noise obtained after filtering the D/A.

Temperature sensitivity At any fixed digital input, the analog output will vary with temperature. This temperature sensitivity typically ranges from about

$\pm 50 \text{ ppm}/^\circ\text{C}$ in a general-purpose converter to as low as $\pm 1.5 \text{ ppm}/^\circ\text{C}$ in a high-quality unit. The overall temperature sensitivity is due to the temperature sensitivities of the reference voltages, the resistors in the converter, the op-amp, and even the amplifier offset voltage.

14.12 A/D CONVERTERS: A PARALLEL-COMPARATOR TYPE

We turn now to systems employed to convert an analog input to a digital output. It should be noted at the outset that generally A/D converters are complex, sophisticated systems. Hence, unlike the situation which prevails with D/A converters, it is ordinarily not feasible to present a detailed schematic diagram. Instead we shall discuss A/D converters in terms of the basic digital and analog building blocks we have already described. These basic blocks include gates, flip-flops, registers, counters, D/A converters, comparators, etc. It is further to be noted that there are many possible schemes for A/D conversion and many variations possible within each scheme. The very large number of A/D converters described in the literature (a good number of which are available commercially) is a tribute to the persistence and ingenuity of engineers who have applied themselves to the problem. In the present and succeeding sections we shall describe a number of representative systems.

In the comparator A/D converter shown in Fig. 14.12-1 the range of analog input extends from 0 to V_0 , and a 3-bit digital output is provided.

The relationship of the digital output to the analog input is presented in Fig. 14.12-2. The analog input is divided into eight ranges. Six of these ranges encompass an interval $S = V_0/7$. The other two ranges, at the ends, extend over the interval $S/2 = V_0/14$. When the analog input is anywhere in the lowest range from 0 to $V_0/14$, the A/D converter output is to be 000, as indicated. As also indicated, if this digital output were in turn to be reconverted to an analog voltage, as by a D/A converter, the analog reading would be 0 V. Hence, in the A/D conversion, an error, the *quantization error*, has been introduced. In this lowest range the error is at most equal to $S/2 = V_0/14$. Similarly when the input is in the range S extending from $V_0/14$ to $(3 V_0)/14$ the corresponding digital output will be 001. This 001 will be interpreted as representing the analog voltage $V_0/7 = 2V_0/14$. Hence in this interval the quantization error will again never be larger than $S/2 = V_0/14$ no matter where in the range the input falls. It can now be seen that the reason for setting the intervals as in Fig. 14.12-2 is to arrange that throughout the input range from 0 to V_0 the maximum quantization error will be the same.

It will be recalled that a comparator (see Sec. 2.13) is a device with two inputs (a reference input and a signal input) and a single output. When the signal input is less than the reference-input voltage, the comparator output is at logic 0. When the signal is higher than the reference, the output is at logic 1. A comparator A/D converter with N output bits requires $2^N - 1$ comparators. For the 3-bit system of Fig. 14.12-1, seven comparators C_1 to C_7 are used.

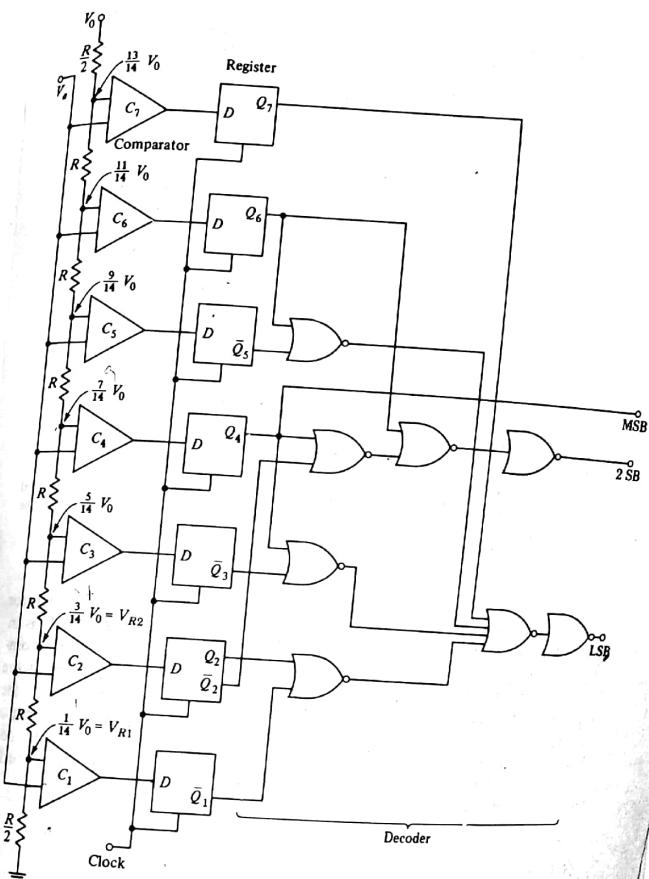


FIGURE 14.12-1
A comparator A/D converter.

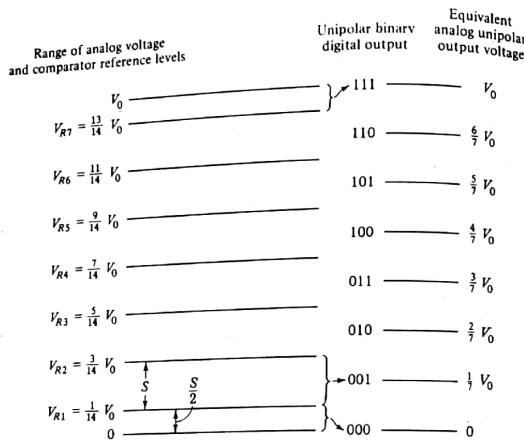


FIGURE 14.12-2
A unipolar analog voltage is divided into intervals and the intervals assigned digital representations in a manner assuring uniform maximum quantization error over the entire range.

If the analog voltage V_a is in the range 0 to $\frac{1}{14}V_0$, all the comparator output logic levels will be 0, that is, $C_1C_2C_3C_4C_5C_6C_7 = 0000000$. If V_a is in the range $\frac{1}{14}V_0$ to $\frac{2}{14}V_0$, then $C_1C_2C_3C_4C_5C_6C_7 = 1000000$, etc. These comparator outputs will be transferred to the outputs of the seven flip-flops of the register at the occurrence of a clock pulse. Finally, as in Fig. 14.12-1, the register is followed by a decoder, which converts the register indications into a 3-bit unipolar binary code. It can be verified (see Prob. 14.12-1) that the decoder indicated in Fig. 14.12-1 does indeed assign the 3-bit output code words in the manner required in Fig. 14.12-2.

EXAMPLE 14.12-1 If the comparator A/D converter shown in Fig. 14.12-1 is to convert analog voltages varying from $-V_0$ to $+V_0$ into two's-complement arithmetic, determine the reference voltages at each comparator input. Illustrate the result using a chart similar to Fig. 14.12-2. Assume a 3-bit output.

SOLUTION When the output of the A/D converter is 000, it will be read as 0 V. Hence, if the maximum quantization error is to be $S/2$, as in Fig. 14.12-2, the output indication 000 should be assigned to the analog range $0 \text{ V} \pm S/2$, as shown in Fig. 14.12-3. A slight difficulty now arises because the analog range is symmetrical about 0 V

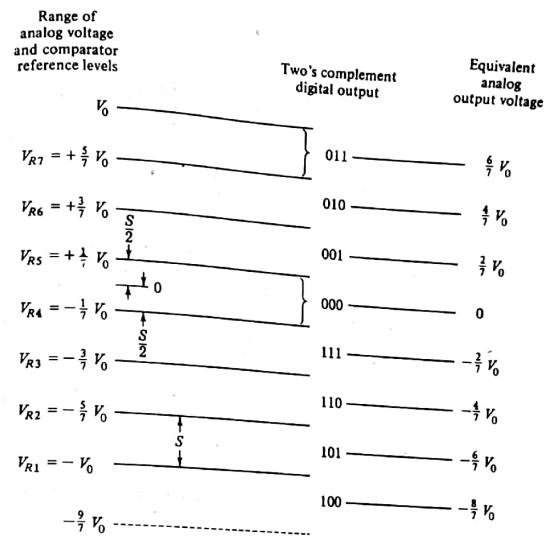


FIGURE 14.12-3
Reference levels and digital outputs for the comparator A/D converter using the two's-complement representation.

while in the two's-complement representation there is always one more negative number than positive number (see Sec. 11.9). If we ignore for the moment the most negative digital number ($100 = -4$), seven digital outputs remain. Accordingly the analog range V_0 to $-V_0$ is divided into seven intervals, each of size $S = 2V_0/7$, and, as shown in Fig. 14.12-3, one digital output is assigned to each range.

We still have the extra digital output 100, which will serve to represent the range $-\frac{9}{7}V_0 \pm S/2 = -\frac{9}{7}V_0 \pm \frac{1}{7}V_0$, that is, the range $-V_0$ to $-\frac{8}{7}V_0$. If we choose to use this output, the bottom of the resistor chain in Fig. 14.12-3 will have to be returned to $-9V_0/7$ and seven comparators will be required with reference voltages $-V_0$, $-\frac{9}{7}V_0$, etc., up to $\frac{1}{7}V_0$. If we choose to ignore this 100 output, the resistor chain will be returned to $-V_0$ and only six comparators will be needed. Most manufacturers choose to use the 100 output and specify their device as having a corresponding asymmetrical voltage range.

The comparator A/D converter is capable of great speed since the entire conversion process occurs simultaneously rather than sequentially. Its operation is fast enough to warrant the use of emitter-coupled logic for the flip-flops

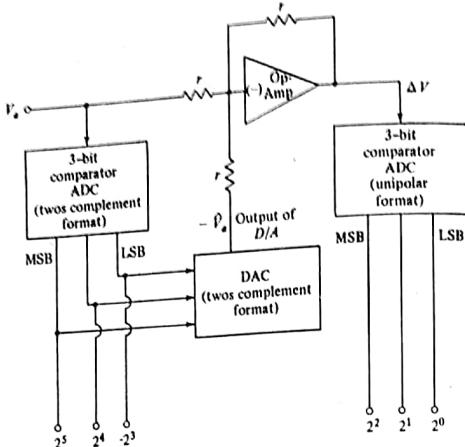


FIGURE 14.12-4
A 6-bit A/D converter comprising a cascade of two 3-bit comparator A/D converters.

and the decoder gates. Immediately on presentation of the analog input voltage and after only the short delay in the comparators, a digital representation of the analog voltage is available. We may note that the flip-flop register is not essential to the operation. However, the register is convenient, especially when the analog input is changing rapidly. The register permits us to hold the digital representation of the analog input until we are ready to accept a new sample.

A cascade-comparator A/D converter The comparator A/D converter has an inconvenient feature in that the hardware required nominally doubles for each additional output bit. Thus, while a 3-bit A/D converter requires seven comparators and seven flip-flops, a 4-bit A/D converter would require fifteen comparators, fifteen flip-flops, and a corresponding increase in decoding gates. At a sacrifice of speed of operation, it is possible to cascade comparator converters with a saving in hardware. Such a cascaded converter is shown in Fig. 14.12-4. The converter provides 6 output bits. Such a converter, if constructed as a single unit after the pattern of Fig. 14.12-1, would require $2^6 - 1 = 63$ comparators. In Fig. 14.12-4 however, we use two 3-bit converters involving $2(2^3 - 1) = 14$ comparators.

The first converter ADC-1 provides the three most significant bits while ADC-2 generates the three least significant bits. Let us assume for convenience that the scaling of the overall 6-bit converter is arranged so that the output

reads directly in volts, (e.g., 001101 represents 13 V). In this case the step size of ADC-2 is $S_2 = 1$ V while the step size of ADC-1 is 8 V. The digital output of ADC-1 is applied to a D/A converter which yields an analog output V_d . The analog voltage difference is converted to digital form by the second converter ADC-2.

We note, most importantly, that the bits added by ADC-2 can only leave unchanged or increase the numerical significance of the final 6-bit representation. It would then appear that we must arrange that ΔV always be zero or positive. We may assure that such be the case by setting the comparator reference levels in ADC-1 at ..., -8 V, 0 V, 8 V, 16 V, Thereafter we shall decode in such manner that when V_a is in the range -8 to 0 V the digital output be 111 (using two's complement representation), when V_a is in the range 0 to 8 V the output be 000, when V_a is in the range 8 to 16 V the output be 001, etc. Such setting of comparator levels and subsequent decoding would be nearly but not quite what is required. For, consider that V_a is infinitesimally smaller than 8 V. We shall require in this case that the 6-bit output read 001000 in order that the quantization error be no more than $\pm \frac{1}{2}$ LSB which in the present case is ± 0.5 V. However, with V_a just less than 8 V, the digital output of ADC-1 would be 000. Even if ADC-2 yielded an output 111 the 6-bit indication would be 000111 with an error of a full LSB. As is readily verified, the difficulty may be remedied by moving the comparator reference levels in the negative direction by $\frac{1}{2}$ LSB = $\frac{1}{2} S_2 = 0.5$ V. In this case ΔV may turn out to be difficult results.

The difference ΔV applied to ADC-2 will now be in the range from -0.5 to 7.5 V. The unipolar format of comparator reference levels as displayed in Fig. 14.12-2 may be set to accommodate this range with quantization error ± 0.5 V. To accomplish this end we need only to set the comparator reference levels at 0.5, 1.5, ..., 6.5 V.

Sources of error A quantization error is necessarily introduced when an analog signal is converted to digital form. This error is often specified as $\pm \frac{1}{2}$ LSB; that is, it has a magnitude which is one-half the numerical significance of the least significant bit of the digital output. For example, consider a straightforward (single-stage) converter intended to handle an input analog signal in the range 0 to 7 V with a 3-bit digital output from 000 to 111. The reference level of the seven comparators would be set at 0.5, 1.5, ..., 5.5, 6.5 V. If, say, the analog voltage V were in the range $2.5 < V < 3.5$, then the converter output would read 011 = 3.0 V. In a physical situation, however, the error might be larger than $\pm \frac{1}{2}$ LSB. Such would be the case if the comparator reference levels were not precisely set and if the comparators operated less than ideally. Ideal operation of the comparator would require the comparator outputs to be at logic 1 or at logic 0 as the analog voltage is infinitesimally on one side or the other of the reference level, there being no range of uncertainty.

In the cascaded comparator converter we have sources of error not only in the comparators but also in the D/A converter, the mechanism used to form the difference $V - V_s$, and the amplifier. These additional operations, none of which is present in the straightforward comparator converter, must be interposed between every stage of a cascaded converter.

High-speed operation Although the cascaded A/D converter serves significantly to reduce the number of comparators required in the system, it also reduces the speed of operation by a factor of 2. For on the first clock pulse we convert the first 3 bits, and on the second clock pulse we convert the second 3 bits.

With some extra circuitry, the cascaded A/D converter can be operated at the same speed as the A/D converter shown in Fig. 14.12-1. We note that while the least significant 3 bits are being converted, the A/D converter for the most significant 3 bits is not performing any useful function. Thus, we can design circuits which enable this first converter to start operating on the next sample of V_s .

14.13 SUCCESSIVE-APPROXIMATION CONVERTER

The principle of the successive-approximation converter is set forth by the following example. Suppose that we have an object whose weight is unknown beyond the fact that it is in the range 0 to 1 kg. Suppose further that a balance is available and a set of known weights of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ kg, etc. These known weights are to be used in a succession of trials to determine the unknown weight. With the unknown weight W on one side of the balance we place the $\frac{1}{2}$ kg on the other side. If we find $W > \frac{1}{2}$ kg, we leave the $\frac{1}{2}$ -kg weight on the scale and add the $\frac{1}{4}$ -kg weight. If we find $W < \frac{1}{2}$ kg, we remove the $\frac{1}{4}$ -kg weight and put on the $\frac{1}{8}$ -kg weight. In this way we continue to try weights successively smaller by a factor of 2. Whenever the last trial weight added tilts the balance toward the side of the known trial weights, we remove the last weight added and try the next smaller weight. Thus, if we found that we could leave the $\frac{1}{2}$ -kg weight, had to remove the $\frac{1}{4}$ -kg weight, and could leave the $\frac{1}{8}$ -kg weight, we would approximate the unknown weight as $1 \times \frac{1}{2} \text{ kg} + 0 \times \frac{1}{4} \text{ kg} + 1 \times \frac{1}{8} \text{ kg} = \frac{5}{8} \text{ kg}$. Assigning the numerical significance $\frac{1}{2}$ to the most significant binary digit, $\frac{1}{4}$ to the next, etc., we would have for the weight the binary designation 101. It is clear that continuing this operation with successively smaller weights, we can establish the unknown weight to whatever precision we please.

If the number of allowable successive weighings is limitless, the procedure described above is acceptable. Suppose, however (as is normally the case), that the number of weighings is finite. Then for the sake of reducing the quantization error it is necessary to offset the scale, i.e., to bias or "tilt" the scale in favor of the unknown. The magnitude of the offset must be equal one-half the smallest weight. That such is the case is illustrated in Fig. 14.13-1. Here we assume two successive trials using weights of $\frac{1}{2}$ and $\frac{1}{4}$ kg to determine a

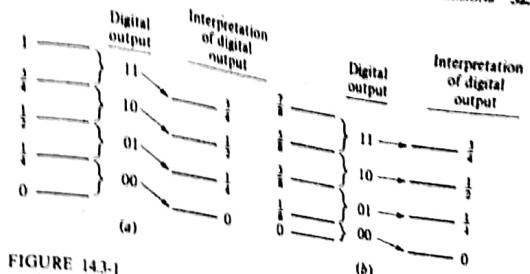


FIGURE 14.13-1
In successive approximation weighing it is necessary to offset the scale by one-half the smallest trial weight.

weight assumed to lie in the range 0 to 1. In Fig. 14.13-1a, the range 0 to 1 is divided into four intervals, and each interval is identified by its appropriate digital representation. Also indicated is the interpretation given to each digital representation. Now suppose that the unknown is infinitesimally smaller than $\frac{1}{4}$. We would find that we could use neither the $\frac{1}{2}$ nor the $\frac{1}{4}$ trial weights and the corresponding digital indication would be 00. The quantization error is then $\frac{1}{16}$. A similar maximum quantization error of $\frac{1}{16}$ will be encountered in every other interval.

Now suppose we offset the scale by adding a weight of $\frac{1}{4}$ to the scale side intended for the unknown. Then the four digital representations will be associated with weight ranges in the manner shown in Fig. 14.13-1b. An unknown weight just less than $\frac{1}{4}$ will appear as a weight just less than $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. This weight falls in the range with digital indication 01. This indication will be interpreted as $\frac{1}{8}$ so that the quantization error is $\frac{1}{16}$. As can be seen in Fig. 14.13-1b, the new range of the unknown is from 0 to $\frac{7}{16}$, and, as can be verified, any place in this range the maximum quantization error to be encountered is $\frac{1}{16}$.

A 3-bit successive-approximation A/D converter A diagram of a 3-bit successive-approximation A/D converter is shown in Fig. 14.13-2. This converter is designed to convert an analog waveform into binary code neglecting the sign bit. For simplicity we have not been as economical of hardware as we might otherwise have been. In this A/D converter we have allowed five (equal) time intervals to accomplish a single A/D conversion. Three of these intervals are used to determine the 3 digital bits; a fourth interval is used simply to read the digital output; while a fifth interval is used to clear the converter in readiness for the next conversion.

The five type-D flip-flops FFA to FFE are connected to form a modulo-5 ring counter (see Sec. 10.16). Such a counter provides at its outputs Q_4 to Q_0 , five waveforms, only one of which is at logic level 1 at any time; the logical level 1 is transferred from A to B to C, etc., with each successive clock cycle.

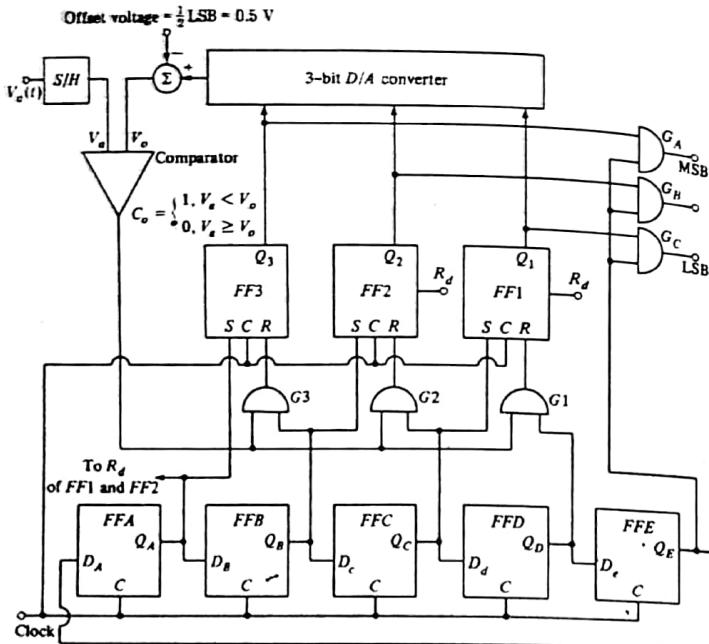


FIGURE 14.13-2
A 3-bit successive-approximation A/D converter.

The three flip-flops $FF1$, $FF2$, and $FF3$ are employed to register the digital bits, with $FF1$ corresponding to the LSB and $FF3$ to the MSB.

The conversion cycle begins with $Q_4 = 1$, while $Q_5 = Q_6 = Q_7 = Q_8 = 0$. Then $FF3$ will be set while $FF2$ and $FF1$ will be reset. We then have $Q_3 = 1$ and $Q_2 = Q_1 = 0$. The input 100 is thus presented to the 3-bit D/A converter, which provides a corresponding analog output V_o . The comparator output C_o will then be $C_o = 0$ or 1 depending on whether $V_a \geq V_o$ or $V_a < V_o$. During the next clock interval $Q_4 = 1$, while $Q_4 = Q_5 = Q_6 = Q_7 = 0$. With $Q_4 = 1$ the AND gate $G3$ is enabled, and $FF3$ is reset if $C_o = 1$ and left in the set state if $C_o = 0$. Thus, altogether we have tentatively assigned a logic 1 to the most significant position, and at the beginning of the second clock interval this bit remains or is changed to logic 0 depending on the comparison of V_a and V_o .

During succeeding clock intervals the trial is repeated for the bits in the next two places. The interval when $Q_4 = 1$ is an interval when no comparisons are being made and we can read the digital output. Thus, Q_4 is used to strobe the output gates G_4 , G_5 , and G_6 .

The point need hardly be belabored that during the sequence of operations leading finally to a digital output the sample value of the analog input must be held constant. Therefore, the converter of Fig. 14.13-2 must be preceded by an S/H amplifier. The sample operation must be synchronized to the operation of the converter. The read-out interval when $Q_4 = 1$ is an interval suitable for sampling, while during the interval from $Q_4 = 1$ through $Q_6 = 1$, that is, while $Q_4 = 0$, the sampled signal is held. Hence, the synchronization can be effected by using Q_4 to operate the switches in the S/H circuit.

As the converter operates, the flip-flops must toggle back and forth, the D/A switches must open and close, and the comparator sees an input which switches abruptly from one level to another. There are transients associated with all these switchings and abrupt changes, and the speed at which the converter can be driven depends on the decay time of these transients. However, the accuracy of the converter does not depend on the flip-flop or on the gates. The accuracy depends almost exclusively on the accuracy of the D/A converter and hence on the accuracy of the D/A resistors, etc.

At the outset of our discussion of the successive-approximation method of conversion we described the method in terms of a succession of weight comparisons on a scale. Returning to that analogy, it is seen that the comparator shown in Fig. 14.13-2 is the scale, the analog input V_a is the unknown weight, and the A/D converter output V_o is the sum of all the trial weights which are "left on the scale." Whenever a bit 1 causes the scale to tilt toward the side of the sum of the trial weights ($V_o > V_a$), the bit 1 is reset to a 0 and the next bit of lower significance is tried.

We noted earlier the need to tilt the scale in the direction of the unknown. In the circuit of Fig. 14.13-2 this tilt can be effected by adding a fixed voltage to the unknown V_a or, generally more conveniently, by offsetting V_o in the opposite direction. (One arrangement for providing such an offset is shown in Fig. 14.10-1.) Consider by way of example a 3-bit converter with outputs ranging from 000 to 111 intended for an analog voltage in the range 0 to 7 V. The least significant bit corresponds to an analog voltage of 1 V. Hence the required offset of V_o is $\frac{1}{2}$ V. Specifically the D/A converter must be offset such that when its digital input is 000, its analog output must be -0.5 V, as shown in Fig. 14.13-3. In this case the maximum quantization error will also be 0.5 V. Further, the maximum analog voltage which can be accommodated is 7.5 V. For when the output is 111, the quantization error does not exceed 0.5 V until $V_o > 7.5$ V.

It is instructive to trace the steps needed to encode an analog voltage V_a . If, say, $V_a = 4.9$ V, we begin with $Q_4 = 1$ and $Q_3 Q_2 Q_1 = 100$ so that V_o reads $4 - 0.5 = 3.5$ V. Since $V_a > V_o$, when Q_4 becomes 1, Q_3 remains 1 and Q_2 also

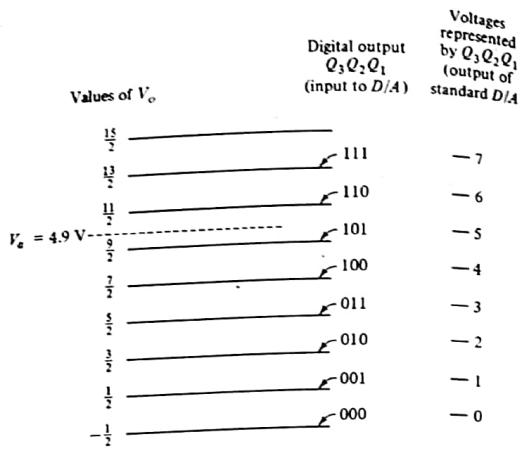


FIGURE 14.13-3
An offset voltage is necessary in the D/A converter of a successive approximation A/D converter.

becomes 1 so that $Q_3 Q_2 Q_1 = 110$. Now $V_o = 6 - 0.5 = 5.5 \text{ V} > V_a$. Hence, when $Q_c = 1$, Q_2 becomes 0 and Q_1 goes to 1 yielding $Q_3 Q_2 Q_1 = 101$. The result is $V_o = 5 - 0.5 = 4.5 \text{ V} < V_a$. Therefore Q_1 does not change when $Q_d = 1$, and the answer is $Q_3 Q_2 Q_1 = 101$. This result is read when $Q_E = 1$.

14.14 THE COUNTING CONVERTER

A three-bit counting converter is shown in Fig. 14.14-1. The three flip-flops are connected as a modulo-8 ripple counter. The counter goes from state to state at each clock pulse when gate G_0 is enabled. Let us assume initially that the control line H (hold count) is at logic level 1. Then G_0 is disabled and the counting does not proceed. Let us assume further that the R_d (reset) line has been used to reset the counter to 000, in which case the output will also read 000. Finally, let us consider that the H line is also used to control the operation of the S/H circuit, so that the waveform $V_a(t)$ is sampled when $H = 1$ and is held when $H = 0$.

Initially, let $V_o = \frac{1}{2}$ LSB and let the comparator output be at logic 1. Now let H change to $H = 0$. Then G_0 is enabled, and the counting proceeds. With each clock pulse the counter advances one count and the D/A converter output V_o jumps by one step. Eventually, we shall have $V_o > V_a$. At that time the comparator output C_o will become $C_o = 0$, and the counting will stop. The count which has accumulated in the counter is the digital output and will be proportional (except for a quantization error) to the analog voltage $V_a(t)$. Having allowed a counting time long enough to ensure that V_o has exceeded V_a , we can

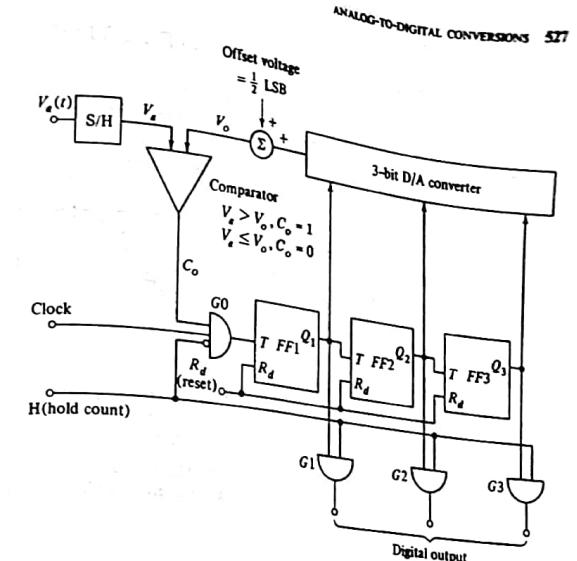
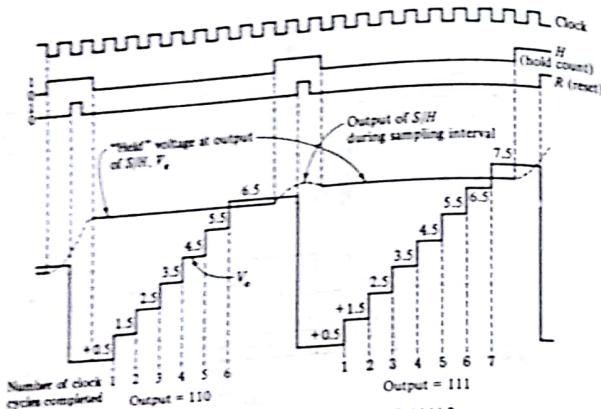


FIGURE 14.14-1
A 3-bit counting A/D converter.

raise H to $H = 1$, thereby allowing a reading of the digital output and allowing the S/H circuit to sample the input signal again. Before returning H to $H = 0$ we move the reset R_d briefly to $R_d = 1$ to clear the counter and then return R_d to $R_d = 0$. Now returning H to $H = 0$ will start a new conversion cycle.

In the counting A/D converter, as in the successive-approximation converter, it is necessary to offset the output V_o of the D/A converter. In the present case, however, the offset must be in the direction to increase rather than to decrease V_o by the voltage corresponding to $\frac{1}{2}$ LSB. To see that such is the case neglect the offset voltage and let the LSB equal 1.0 V, in which case the maximum quantization error is to be ± 0.5 V. Suppose now that V_o is infinitesimally higher than 0 V. Then at the start of the conversion we shall have $C_o = 1$, and the counter will advance by one count and then stop. But this one count will yield a digital output 001 with the interpretation 1.0 V. Correspondingly the quantization error will be 1.0 V. On the other hand, with an offset of 0.5 V the counter will not advance by one count until $V_o > 0.5$ V.

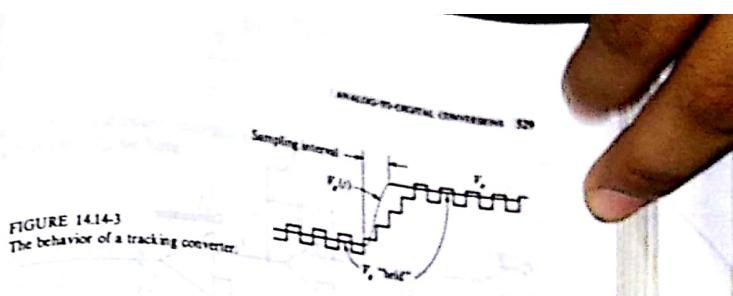
The converter waveforms are shown in Fig. 14.14-2. We assume as before a D/A converter which yields outputs (exclusive of the offset) of 0 and 7 V for digital inputs 000 and 111, respectively. Because of the appearance of the waveform V_o , the counter converter is also referred to as the *digital-ramp converter*. Two conversion intervals are shown. In the first, the analog voltage

FIGURE 14.14-2
A/D converter waveforms for Fig. 14.13-1.

lies between 5.5 and 6.5 V, and the digital output is 110 = 6 V. In the second, V_o lies between 6.5 and 7.5, and the output is 111. The largest analog input which can be accommodated with maximum quantization error 0.5 V is again 7.5 V.

For a given sampling rate and number of output bits the counter converter generally requires a much faster clock rate than the successive-approximation converter. In the counter converter, with N output bits, 2^N clock cycles are required for a conversion operation. In the successive-approximation converter the number required is N (or $N + 2$ if we include clock intervals for resetting and reading, as in Fig. 14.13-2). In any event the clock frequency increases exponentially with N in the counter converter and only linearly in the successive-approximation counter. The counter type of A/D converter is usually restricted to sampling frequencies which are less than 100 kHz, while with successive-approximation converters 1-MHz sampling rates are feasible.

At the expense of some increase in complexity the counter converter can be improved by substituting for the up counter of Fig. 14.14-1 an up-down counter (see Sec. 10.14). Such a converter is referred to as a *continuous-digital-ramp converter*, a *tracking converter*, or a *servo converter*. The counter is commanded to count up or down depending on whether the output of the comparator is at logic 1 or logic 0 and hence depending on whether V_o is larger or smaller than V_s . If initially $V_o > V_s$, the counter counts up until $V_o > V_s$. At this point the counter reverses. If, after a single count down we find that $V_o < V_s$, the counter reverses again, and so on. The level of V_o will therefore

FIGURE 14.14-3
The behavior of a tracking converter.

hunt back and forth across V_s . The appearance of V_o during intervals when V_s is being held and during an intervening sampling interval is indicated in Fig. 14.14-3. The output of this converter is read, as before, at the end of the hold interval. Compared with the straight counting converter, in the servo converter, only half as many counts on the average will be required to complete a conversion. Hence a servo converter can operate at twice the speed.

Of the three types of converters just described which involve feedback through a D/A converter, the successive-approximation converter is by far the most popular.

14.15 THE DUAL-SLOPE CONVERTER

We consider now a type of converter which involves no feedback. This converter, shown in Fig. 14.15-1a and called the *dual-slope converter*, is often used in digital voltmeters. We now describe its principle of operation.

At the beginning of the conversion process, say at $t = 0$, the switch S_1 is connected to point A , and the held sample of the analog input V_s is applied to the analog integrator. If $\tau = RC$ is the time constant of the integrator, the integrator output is $V_i = -(t/\tau)V_s$. The waveform of V_i is shown in Fig. 14.15-1b. At the same time ($t = 0$) a clock waveform is applied to a counter which was initially clear. The counter counts until the counter flip-flops FF0 to FF($N - 1$) simultaneously reset, so that $Q_0 = Q_1 = Q_{N-1} = 0$, at which time FFN is set, that is, $Q_N = 1$. This output Q_N controls the state of S_1 , and when $Q_N = 1$, switch S_1 moves to point B . The output of the integrator now starts to move in the positive direction since the applied reference voltage is negative, that is, $-V_r$ (see Fig. 14.15-1b). The counter continues to count until the output voltage V_o becomes just barely positive. At this time the comparator output goes to the 0 state, gate G_1 is disabled, and the counting stops.

We now show that the count recorded in the N -stage counter, Q_{N-1}, \dots, Q_1, Q_0 , is directly proportional to V_s and is independent of the time constant τ . The time T_1 required for the $N + 1$ flip-flops to go from $00 \dots 00$ to $10 \dots 00$ is $2^N T_c$, where T_c is the time between clock pulses. At this time the output voltage V_o is

$$V_o = -\frac{V_s}{\tau} T_1 = -\frac{V_s}{\tau} 2^N T_c \quad (14.15-1)$$

If at time T_2 the count recorded in the first N flip-flops is λ , since the count was 0 at time T_1 , we have

$$T_2 - T_1 = \lambda T_c = \frac{V_a}{V_r} 2^N T_c \quad (14.15-4)$$

so that the count λ is

$$\lambda = \frac{V_a}{V_r} 2^N \quad (14.15-5)$$

As long as $V_a < V_r$, the system operates as an A/D converter. Since $\lambda < 2^N$, the count is directly proportional to V_a and is a number which can be read from the counter. The converter can be made direct-reading if $V_r = 2^N$ V. Now $\lambda = V_a$, and the count recorded in the counter is numerically equal to the applied voltage V_a .

When the counting ceases, the outputs of $FF1$ to $FF(N-1)$ are recorded, all $N+1$ flip-flops are reset, and capacitor C is discharged using switch S_2 . A new sample can now be converted. The logic circuitry needed to reset the system is left as a problem (Prob. 14.15-2).

14.16 A COMPARISON OF CONVERTER TYPES

The converters we have discussed represent devices whose speed of operation lies in three different ranges. Fastest is the comparator converter. In principle, except for the delay through the comparators, this converter makes available a digital output at the moment the analog input is applied. Hence, this converter is the system of choice where maximum speed is required. If the hardware requirements of a straightforward comparator converter become excessive, a cascaded arrangement can be used with some sacrifice in speed and accuracy.

Next in order of speed is the successive-approximation converter. Where a relatively fast converter of good quality is required, this comparator is by far the most popular. As we have noted, the time required to process a conversion increases linearly with the number of bits, requiring about as many clock pulses as bits. Counter converters are the slowest, requiring 2^N clock cycles per conversion, N being the number of bits. Very popular among counter converters is the dual-slope converter, which is widely used in such instruments as digital voltmeters, where conversion speed is not important.

Beyond the comparators we have discussed there are almost limitless other types, some differing in principle and some only in detail. We consider now briefly and in no great detail some additional converters which have the merit of simplicity and economy.

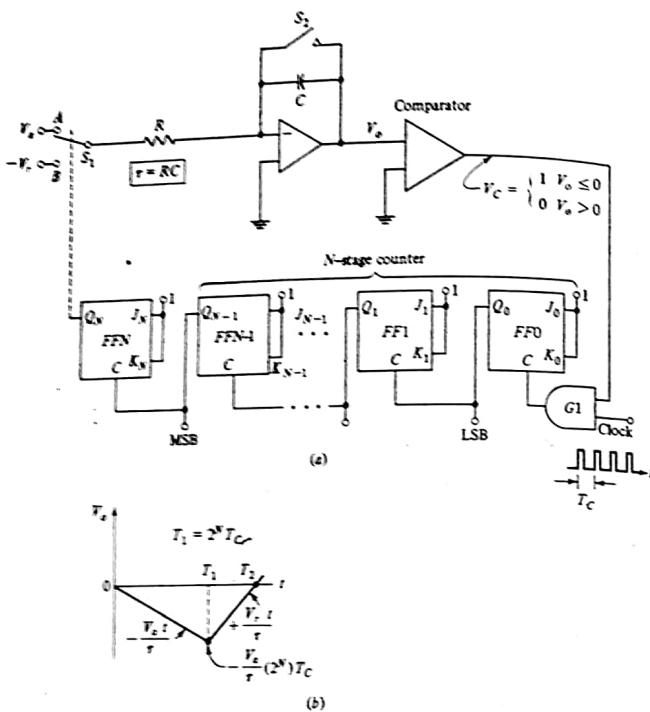


FIGURE 14.15-1
The dual-slope A/D converter.

Referring to Fig. 14.15-1b, we see that at the time T_1 V_o is again equal to 0 V and that therefore

$$\frac{V_a(T_2 - T_1)}{\tau} = \frac{V_a}{V_r} T_1 \quad (14.15-2)$$

Hence, the time interval $T_2 - T_1$ is

$$T_2 - T_1 = \frac{V_a}{V_r} 2^N T_c \quad (14.15-3)$$

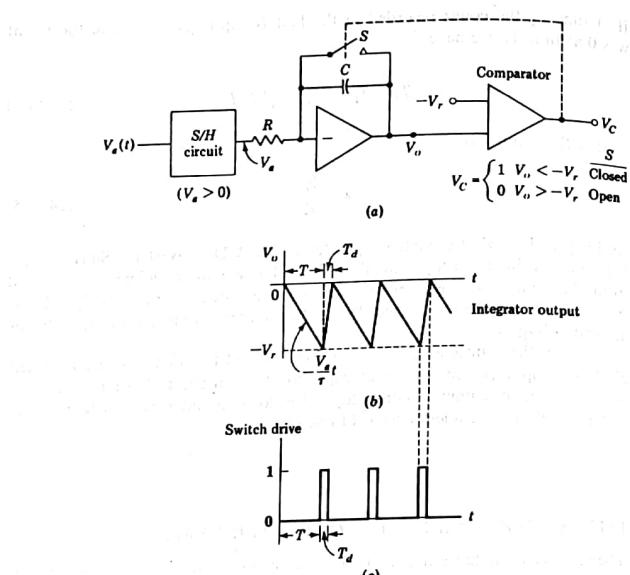


FIGURE 14.17-1
(a) A V/F converter. (b) The waveform V_o . (c) The waveform which drives the switch.

14.17 A CONVERTER USING VOLTAGE-TO-FREQUENCY CONVERSION

An A/D converter can be built using a counter and a device referred to as a *voltage-to-frequency* (*V/F*) converter. In rather simplified form, the *V/F* converter is shown in Fig. 14.17-1a. The principles of operation of this circuit are explained below.

An analog waveform $V_a(t)$ (assumed positive) is sampled and held to form the voltage V_s . This voltage is applied to an integrator, which is followed by a comparator. The other input to the comparator is a reference voltage $-V_r$. Initially the switch S bridging the integrating capacitor C is open, and the voltage V_o decreases linearly with time. If $\tau = RC$, we have $V_o = -V_a t / \tau$, which is shown in Fig. 14.17-1b. When V_o decreases to $-V_r$, after a time $t = T$, the comparator output V_c becomes positive for a small time interval T_d , during which time switch S closes, thereby discharging capacitor C and returning the

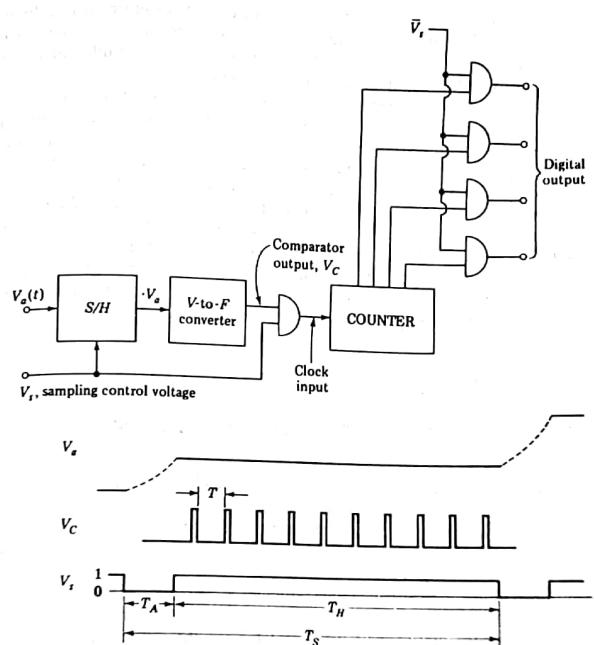


FIGURE 14.17-2
An A/D converter using a *V/F* converter and a counter.

integrator output V_o to approximately 0 V. The discharge rate is, of course, determined by the switch resistance. In an actual system, the comparator output might remain positive for too short a time to discharge the capacitor completely. In this case, a circuit called a monostable multivibrator (see Secs. 15.3-15.5), capable of using the narrow comparator pulse to form a pulse of width T_d , may be interposed between the comparator output and the switch.

After the time interval T_d the comparator voltage drops to the 0 state, switch S opens, and V_o starts to decrease once more. If the discharge time T_d is much less than the integration time T , the frequency of the waveforms V_o and V_c is

$$f = \frac{1}{T + T_d} \approx \frac{1}{T} = \frac{1}{\tau} \frac{V_s}{V_a} \quad (14.17-1)$$

Thus, the V/F converter makes available an output waveform whose frequency is proportional to the input voltage. Such a waveform is said to be *frequency-modulated*. V/F converters, which operate in a manner similar to the device shown in Fig. 14.17-1a, are commercially available.

An A/D converter using a V/F converter is shown in Fig. 14.17-2. The V/F converter provides the clock input (which is the comparator output V_C) to a counter through an AND gate. A second input to the AND gate is the sampling voltage V_s , which holds the logic level 1 for a fixed time T_H . As long as $V_s = 1$, the sampled output V_a is held fixed at its value at the beginning of the interval T_H . With the AND gate enabled for the time T_H the counter reading will equal the number of cycles executed by the V/F converter output in the specified time interval. If we call the number read by the counter, λ , then, from Eq. (14.17-1), $\lambda = fT_H = T_H V_a / tV_r$, and the counter registration is proportional to V_a . The counter output is read when V_s is in the 0 state. In addition, during this time interval T_C , a new sample value of $V_a(t)$ is established.

14.18 A CONVERTER USING VOLTAGE-TO-TIME CONVERSION

In the previous section we discussed the principle of an A/D converter which operates by counting the cycles of a *variable-frequency* source for a *fixed period*. Alternatively, a converter may operate on the principle of counting the cycles of a *fixed-frequency* source for a *variable period*. Such an A/D converter is shown in some detail in Fig. 14.18-1.

Assuming a positive analog voltage, a negative reference voltage V_r is applied to an integrator, the output of which provides one input of a comparator. The comparator output is at logic level 1 when the integrator output V_i is less than the analog voltage V_a . A fixed-frequency clock V_{CL} is applied to a counter through an AND gate G . The circuit operates by devising to keep this AND gate enabled only during the time beginning at $t = 0$, when $V_i = 0$, to the time $t = T$, when $V_i = V_a$. Since $V_i = V_r/t$,

$$T = \frac{\tau V_a}{V_r} \quad (14.18-1)$$

If f_{CL} is the clock frequency, during the interval T the count N registered will be

$$N = f_{CL} T = \frac{f_{CL}}{V_r} V_a \quad (14.18-2)$$

The count N , as required, is proportional to V_a . Note that here, unlike the situation which prevails in the dual-slope converter, the calibration of the present converter does depend on the clock frequency and on the integrator time constant.

Waveforms for the A/D converter are shown in Fig. 14.18-1b. The

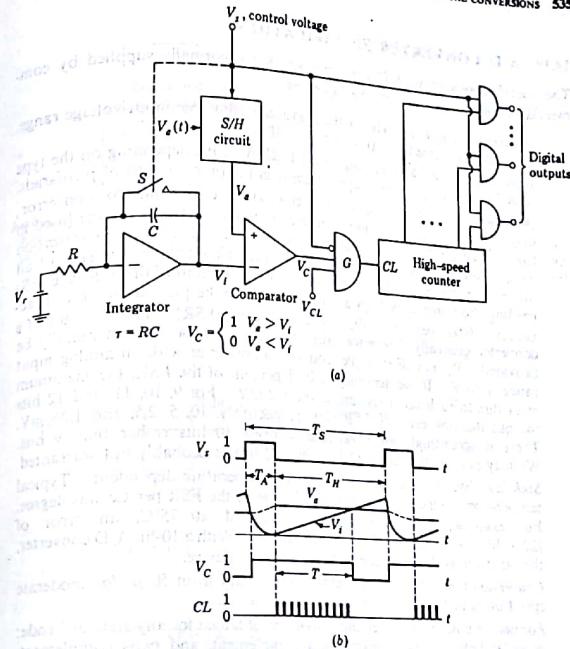


FIGURE 14.18-1

(a) An A/D converter using a voltage-to-time converter. (b) Waveforms.

sampling voltage V_s samples the positive input voltage $V_a(t)$ every T_S during the time interval T_A . The sampled voltage V_a is then held for a time T_H . During this time, switch S is open, and the integrator output is a voltage ramp. When $V_i < V_a$, the comparator output is in the 1 state. Gate G is enabled as long as $V_i < V_a$, i.e., for the time interval T . During the time interval T the clock voltage V_{CL} is transmitted by gate G to the high-speed counter. Thus the counter output is proportional to T . During the time interval T_A , the gate G is disabled and the counter read. In addition, switch S is closed, the capacitor is discharged, and V_i is reset to 0 V.

14.19 A/D CONVERTER SPECIFICATIONS

The specifications of an A/D converter which are normally supplied by commercial manufacturers include the following:

Analog input voltage This is the maximum allowable input-voltage range. Typical values are 0 to 10 V, ± 5 V, ± 10 V, etc.

Input impedance Values range from 1 k Ω to 1 M Ω , depending on the type of A/D converter. Input capacitance is in the range of tens of picofarads.

Accuracy The accuracy of an A/D converter includes quantization error, digital system noise, including that present in the reference voltage (used in the D/A converter), deviations from linearity, etc. Usually the quantization noise is specified as $\pm \frac{1}{2}$ LSB. The accuracy also includes the sum of all other error sources. Typical values are ± 0.02 percent of the full-scale (FS) reading; however, very high accuracy A/Ds can be purchased with an accuracy of 0.001 percent of the full-scale reading (FSR). The accuracy of a converter generally determines the number of bits which can usefully be provided. By way of example, consider a converter with an analog input range ± 10 V. If the accuracy is 0.02 percent of the FSR, the maximum error due to such accuracy limitation is 2 mV. For 9, 10, 11, and 12 bits the quantization error ($\frac{1}{2}$ LSB) are, respectively, 10, 5, 2.5, and 1.25 mV. There is accordingly an advantage in using 10 bits rather than 9 bits. We may even justify using 11 bits, but 12 bits is probably not warranted.

Stability System accuracy is generally temperature-dependent. Typical temperature coefficients of error are 20 ppm of the FSR per Celsius degree. For example, if a 10-V signal is applied at 75°C, an error of $(20 \times 10^{-6})(10\text{ V})(75 - 25) = 10\text{ mV}$ results. With a 10-bit A/D converter, the error limits the response to that of a 9-bit device.

Conversion time Typical conversion times vary from 50 μs for moderate speed units to 50 ns for a very high speed device.

Format An A/D converter can usually be obtained for any standard code: unipolar binary, offset binary, ones complement, and twos complement. In addition the output voltage levels are often adjusted so that direct connection is possible to some logic family (TTL, ECL, etc.).

14.20 INTERCONNECTING THE S/H AND THE A/D CONVERTER

The system employed to convert an analog signal into a digital bit stream consists of a S/H amplifier and an A/D converter. These units are operated in synchronism, the A/D converter "telling" the S/H amplifier when to sample and when to hold.

Figure 14.20-1a shows the S/H-A/D conversion system. Note that the complete A/D converter consists of two subsystems. The first is the A/D converter, such as the comparator converter shown in Fig. 14.12-1 or the successive-approximation converter shown in Fig. 14.13-2, etc. The second subsystem is a *timing circuit*. The input to the timing circuit is the sampling

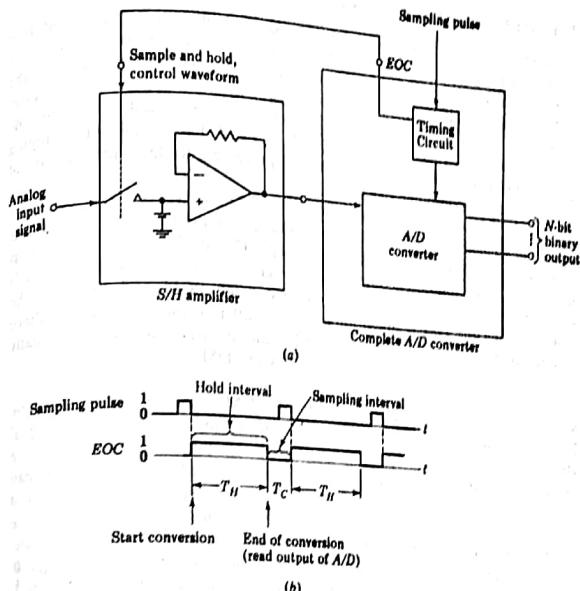


FIGURE 14.20-1
(a) Interconnection of the S/H circuit and A/D converter. (b) Waveforms.

pulse train. The timing circuit provides all the clock pulses needed for the A/D converter to convert the analog sample into an N -bit output signal. In addition the timing circuit generates a timing waveform called the *end-of-conversion (EOC) waveform*, which tells the S/H circuit when to sample and when to hold (see Fig. 14.20-1b).

It is important to note that the sampling pulse train is applied to the A/D converter and not to the S/H circuit. The control employed by the S/H circuit is the EOC waveform generated in the A/D converter. This is done so that the output of the S/H circuit is held constant until the A/D converter has finished converting. Then, while the converted output is displayed at the binary output terminals of the A/D converter, the S/H circuit is permitted to change its analog output level (this occurs during time T_d). If the S/H circuit and A/D converter were not synchronized, one might well envision the S/H output changing during conversion, and such operation would lead to an incorrect digital output.

14.21 DELTA MODULATION¹

The A/D converters discussed previously make available a digital representation of an analog signal. The fundamental characteristic common to each of these devices is that it produced an N -bit code word for each sample taken of the analog signal. Although the type of code may vary (offset binary, ones complement, two's complement, etc.), the generic term employed to describe the operation is *pulse-code modulation* (PCM). PCM has found wide-ranging applications from computer-type arithmetic processing to data transmission of voice, video, and computer signals.

In the following sections we consider a different type of A/D converter called a *delta modulator* (DM). While applications of the DM are not nearly as large as for PCM, they are rapidly increasing. One of the major applications of the DM is the digital encoding of voice signals prior to transmission. It is found that similar clarity of voice, encoded using PCM and DM, can be achieved when the DM is operating at a lower bit rate than for PCM. Since channel bandwidth required increases with bit rate, and since bandwidth is usually at a premium, many digital voice systems now use DM.

A basic "linear" DM. A block diagram of a DM is shown in Fig. 14.21-1a. Typical waveforms are shown in Fig. 14.21-1b, c, and d. The analog signal $M(t)$ and the estimate $\hat{M}(t)$ of $M(t)$ derived by the delta modulator are continually presented to a comparator. The output of the comparator can assume one of two possible voltage levels: logic 1 if $M(t) \geq \hat{M}(t)$ and logic 0 if $M(t) < \hat{M}(t)$. The comparator output waveform (not shown) is sampled once per clock cycle and held at the sampled value for the duration of the clock cycle. If at the sampling time kT_s (k an integer) the comparator output is at logic 0, $E(t) = E(kT_s) = 0$ is transmitted; if, however, the comparator output is at logic 1, $E(t) = E(kT_s) = 1$ is transmitted.

The estimate $\hat{M}(t)$ is formed using the following algorithm: if $E(kT_s) = 1$, which denotes that $M(kT_s) \geq \hat{M}(kT_s)$, increase $\hat{M}(t)$ by a step S_0 during the interval $kT_s + \tau < t < (k+1)T_s + \tau$. If $E(kT_s) = 0$, which denotes that $M(kT_s) < \hat{M}(kT_s)$, decrease $\hat{M}(kT_s)$ by a step S_0 . The time delay τ is the time required for the processor to convert changes in $E(t)$ into changes in $\hat{M}(t)$. The algorithm can be written in terms of the sampling times kT_s and $(k+1)T_s$ as a recursive difference equation

$$M[(k+1)T_s] = \hat{M}(kT_s) + S_0 E(kT_s) \quad \text{for all } k \quad (14.21-1)$$

where $E(kT_s) = \begin{cases} +1 & \text{if comparator output is at logic level 1} \\ -1 & \text{if comparator output is at logic level 0} \end{cases}$

The application of this algorithm results in the estimate shown in Fig. 14.21-1c. The resulting bit stream, $E(t)$, is shown in Fig. 14.21-1d.

One way in which the processor shown in Fig. 14.21-1a can be realized is shown in Fig. 14.21-2. The processor consists of the up-down counter and the D/A converter. When at the sampling time it turns out that $E(kT_s) = 1$, the

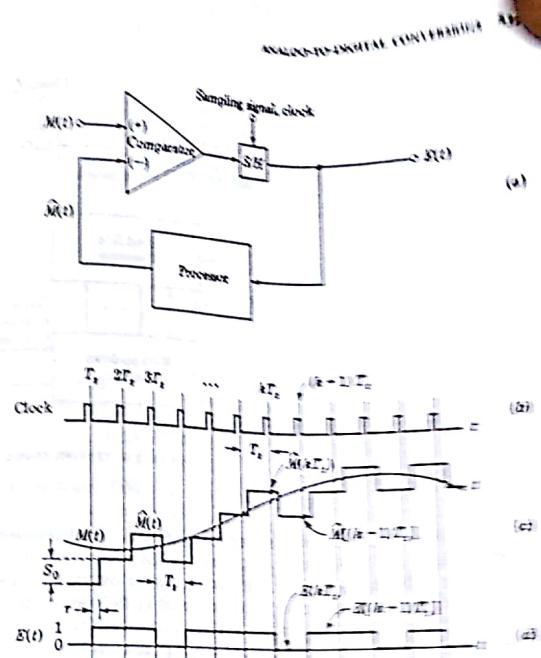


FIGURE 14.21-1
(a) A basic delta modulator. (b) Clock. (c) Signal and estimate. (d) Transmitted bit stream $E(t)$.

number stored in the counter is increased by 1 and when $\hat{M}(kT_s) = 0$, the number stored in the counter is decreased by 1.

Another simple DM is shown in Fig. 14.21-3. Here the processor is not a digital device, consisting rather of an analog integrator. Here $\hat{M}(kT_s) = -V_0/4$ whenever $M(kT_s) < \hat{M}(kT_s)$ and $-V_0/4$ whenever $M(kT_s) > \hat{M}(kT_s)$. In this case the waveform $\hat{M}(t)$ consists not of a sequence of up or down jumps, but a sequence of ramps with slopes $+V_0/4RC$ or $-V_0/4RC$. In the same interval $T_s = 1/f_s$, the total change in $\hat{M}(t)$, which we again call S_0 , is

$$S_0 = \frac{V_0 T_s}{4RC} \quad (14.21-2)$$

Reconstruction of the analog signal One of the great merits of delta modulation is the relative ease with which the quantized analog signal $\hat{M}(t)$ can be

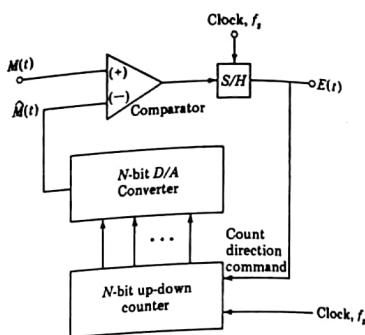


FIGURE 14.21-2
A DM using an up-down counter as a processor.

recovered from the digitized transmitted signal $E(t)$. This recovery process, of course, effects the D/A conversion, which is the inverse of the original A/D conversion. The recovery of $\hat{M}(t)$ is accomplished by passing the transmitted signal $E(t)$ through a processor such as is used at the transmitter. For, as we have noted, the processor generates $\hat{M}(t)$ from the waveform $E(t)$. However, no matter whether the transmitter processor is digital, as in Fig. 14.21-2, or analog, as in Fig. 14.21-3, the simple integrator processor is suitable for signal recovery.

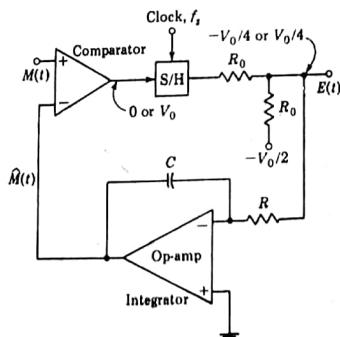


FIGURE 14.21-3
A simple realization of a DM.

Limitation of the linear DM The DM systems described in this section both have the characteristic that the estimate $\hat{M}(t)$ changes by $\pm S_0$, that is, by a constant amount, after each sampling pulse. Thus, if $E(t)$ should consist of a long stream of logic 1 bits, $\hat{M}(t)$ would resemble a voltage ramp or a stepwise approximation to a ramp. As a result of this "linear" change in $\hat{M}(t)$, the type of DM system described above is referred to as *linear delta modulation*.

The linear DM has the fundamental limitation that the magnitude of the "slope" of $\hat{M}(t)$ cannot exceed $S_0/T_s = S_0 f_s$. If then the slope of the signal $M(t)$ should persist at a value greater than S_0/T_s for an extended time, the difference between $M(t)$ and $\hat{M}(t)$ may become quite large. This situation is illustrated in Fig. 14.21-4. The disparity between $M(t)$ and $\hat{M}(t)$ shown here is described as a *slope-overload error*.

To decrease the slope-overload error we can increase the step size S_0 with a corresponding increase in the quantization error, or we can increase the sampling rate f_s with a corresponding increase in bandwidth required to transmit the signal $E(t)$. As a result of this limitation linear DM is rarely used in practice. In the next section we discuss *adaptive DM*, which overcomes the slope-overload problem by generating different step sizes depending on the characteristics of $M(t)$.

14.22 ADAPTIVE DELTA MODULATION

Adaptive DM differs from linear DM in that the step size is no longer constant at S_0 but takes on other values, depending on the past history of the transmitted data, $E(t)$, that is, $E(kT_s), E[(k-1)T_s], \dots, E(0)$. This type of DM is often used to digitally encode voice. The voice reproduction is found to be comparable to that

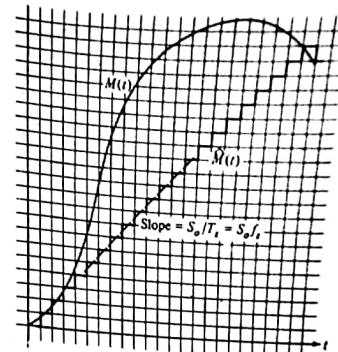


FIGURE 14.21-4
Slope overload.

obtained in PCM, even if the DM is operating at a lower bit rate. (The bit rate in DM is equal to the sampling rate. In PCM the bit rate is equal to the product of the sampling rate and the number of bits into which each sample is encoded.)

In this section we describe an adaptive DM (ADM), designed and constructed at the Communication Systems Laboratory of CCNY²; it can reproduce voice which has been band-limited to 2,400 Hz with a word intelligibility of 90 percent at 9.6 kb/s. This is equivalent to using PCM and encoding each sample into 2 bits.

Quite generally, in DM, we have

$$\hat{M}[(k+1)T_s] = \hat{M}(kT_s) + S[(k+1)T_s] \quad (14.22-1)$$

where $S[(k+1)T_s]$ is the step size by which the estimate $M(t)$ changes at $t = kT_s + \tau$ (see Fig. 14.20-1c). In the linear DM, described above,

$$S[(k+1)T_s] = S_0 E(kT_s) \quad (14.22-2)$$

leading to Eq. (14.21-1). In an ADM system the step size $S[(k+1)T_s]$ is not of constant magnitude, as in Eq. (14.21-2). In the system described here $S[(k+1)T_s]$ is given instead by

$$S[(k+1)T_s] = |S(kT_s)|E(kT_s) + S_0 E[(k-1)T_s] \quad (14.22-3)$$

Shown in Fig. 14.22-1 is a signal $M(t)$ and the estimate $\hat{M}(t)$ which results when $S[(k+1)T_s]$ is generated through the algorithm of Eq. (14.22-3). To verify that $\hat{M}(t)$, as shown, is consistent with Eq. (14.22-3) it is well to keep in mind that $S[(k+1)T_s]$ represents the jump in $\hat{M}(t)$ at the time $kT_s + \tau$, where τ is the processing delay time in the system. However $E(kT_s)$ refers to the processor input voltage just after the k th clock pulse. This is illustrated in Fig. 14.22-1. The signal $M(t)$ in Fig. 14.22-1 is one which would generate a large slope-overload error if linear DM were used. Observe, however, in the present case that as the condition $M(t) > \hat{M}(t)$ persists, the jumps in $\hat{M}(t)$ become progressively larger. The estimate $\hat{M}(t)$ therefore catches up with $M(t)$ sooner than it would with linear modulation. On the other hand, when, in response to a large slope, $M(t)$ develops large jumps, it may well require a large number of clock cycles for these jumps to decay in amplitude when they are no longer required. Such a situation is to be seen in $\hat{M}(t)$ in the neighborhood of the maximum of $M(t)$ in Fig. 14.22-1. Altogether, then, while this ADM system reduces slope error, it does so at the expense of increasing quantization error.

It turns out that in the matter of speech transmission the reduced slope error provides a net advantage in spite of the increased quantization error. The spurious frequency components introduced into the reconstructed signal by slope-overload error are principally in the low-frequency range, while quantization error introduces principally spurious high-frequency components. The power in speech is concentrated largely in the low-frequency components, and these low-frequency components are also the ones which must principally be reproduced without noise contamination in order to preserve intelligibility. Hence, if we

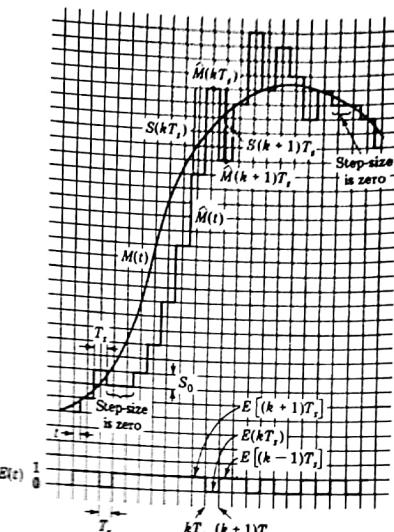


FIGURE 14.22-1
The ability of an ADM to approximate the same waveform as in Fig. 14.20-4. The minimum step size is the same as in Fig. 14.20-4.

pass the reconstructed signal through a low-pass filter, we shall be able to discriminate against the high-frequency noise due to quantization error without materially degrading the voice signal.

It is of incidental interest to note, in connection with $\hat{M}(t)$ in Fig. 14.22-1, that on occasion $\hat{M}(t)$ does not exhibit a jump in each clock interval. Two such instances are indicated by brackets. It can also be verified that in a steady-state condition when $M(t)$ remains constant (within an interval S_0), $\hat{M}(t)$ oscillates about $M(t)$ but the oscillation frequency is one-half the frequency that would be encountered in linear DM.

REFERENCES

1. Taub, H. and D. L. Schilling, "Principles of Communications Systems," McGraw-Hill, New York, 1971.
2. Sung, C. L., J. Gersbach, and D. L. Schilling, "A Robust Delta Modulator," *IEEE Trans. Commun. Technol.*, December 1971.

4 RESISTOR-TRANSISTOR LOGIC (RTL) AND INTEGRATED-INJECTION LOGIC (IIL)

RESISTOR-TRANSISTOR LOGIC (RTL) AND INTEGRATED-INJECTION LOGIC (IIL) 135

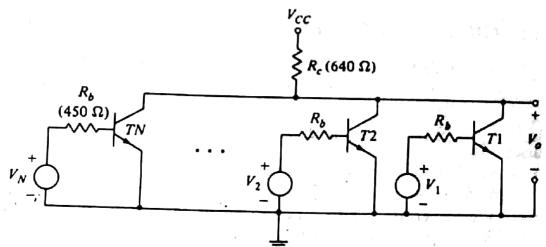


FIGURE 4.1-1
An N -input RTL gate. A medium-power commercial gate has the component values indicated in parentheses.

resistors R_b . If we adopt the convention of *positive logic*, then, as we can now verify, the gate performs NOR logic.

Let V_L and V_H be the voltages representing the two possible values of a logical variable. The subscripts *L* and *H* stand for "low" and "high." That is, $V_L < V_H$. In positive logic V_L corresponds to the case where the logical variable is "false," i.e., to logic level 0. The voltage is V_H when the variable is "true," i.e., at logic level 1. In the RTL gate V_L is required to be low enough for the corresponding transistor to be cut off when V_L is applied at a gate input. The voltage V_H is a voltage which, applied at an input, will drive the corresponding transistor to saturation. Note that V_L and V_H are not unique voltages but are characterized by the fact that V_L must be *less than some unique voltage* and V_H must be *greater than some unique voltage*. In the present case it is necessary that $V_L < V_i$, the cut-in voltage, and V_H be equal to or higher than the voltage which when applied through R_b will bring the transistor to saturation.

The truth table for a two-input gate is given in Table 4.1-1. In one truth table we have used the logical values LOW and HIGH in the other 0 and 1. Both types of tables are common.

Table 4.1-1

V_1	V_2	V_o	V_1	V_2	V_o
L	L	H	0	0	1
L	H	L	0	1	0
H	L	L	1	0	0
H	H	L	1	1	0

We begin our discussion of logic gates (Chaps. 4 to 8) by considering the resistor-transistor-logic (RTL) gate. This type of gate is no longer used in new design, but for a number of reasons the RTL gate is a useful starting point. On the one hand, this gate is elegantly simple and hence may be used conveniently to develop concepts useful in connection with all types of gates. Further, the RTL gate is historically the first gate to have been used extensively, and many installations employing this type of gate are still in operation. Finally, topologically at least, RTL is a forerunner of integrated-injection-logic (IIL) which, at the present writing, is one of the newest of the commercially available LSI logic families.

4.1 THE RESISTOR-TRANSISTOR-LOGIC (RTL) GATE

An N -input RTL gate (Fig. 4.1-1) consists of N transistors all of whose emitters are connected to a common ground and all of whose collectors are tied through a common collector resistor R_c to a supply voltage V_{CC} . Input voltages V_i ($i = 1, 2, \dots, N$) representing logic levels, are applied to the bases through

*Conductors through one of several logic stages becomes
noisier (less) sensitive than others in circuit having*

4.2 THE DIRECT-COUPLED TRANSISTOR-LOGIC (DCTL) GATE

The earliest gates of the type shown in Fig. 4.1-1 were more economically constructed by omitting the base resistors R_b . Because input connections are made directly to the base, such gates are called direct-coupled transistor-logic (DCTL) gates. DCTL gates, like RTL gates, perform NOR logic. We shall discuss these gates briefly calling attention to a fundamental deficiency on account of which they were never widely used.

As noted in Sec. 3.10, the NOR operation is functionally complete. That is, every logic operation can be performed using the NOR operation alone. That is, switching system of any sophistication, then, it is likely that each of the inputs to the NOR gate can be derived from the outputs of other similar NOR gates, and in turn, each NOR gate may be required to furnish input logic levels to other NOR gates. Hence the situation of the DCTL gate G_{11} , shown in the dashed box of Fig. 4.2-1 is representative of the milieu in which a typical gate may have to function.

Let us focus our attention on gate G_{11} . It receives input V_i from gate G_0 , which is also required to furnish a signal to one input of $N - 1$ other gates G_{12} through G_{1N} . Similarly, gate G_{11} must furnish signals to one input of N other gates G_{21} through G_{2N} . The other gate input terminals, not shown connected in Fig. 4.2-1, will derive their inputs from sources not specified here. If any of these gates has more inputs than required for its function, the unused gate input terminals are to be grounded. The corresponding transistor will then be at cutoff and will have no effect on the operation of that gate.

We now determine the voltage levels present at the input and output of the DCTL gates shown in Fig. 4.2-1. Suppose that all inputs of G_0 are low enough to keep all transistors in G_0 cut off. Then the common collector of G_0 will rise, driving transistors T_1, T_2, \dots, T_N to saturation. As noted earlier (Sec. 1.1), at room temperature the base-emitter voltage V_o of a saturated transistor is 0.75 V. Thus the gate voltage level corresponding to the 1 state at the output of gate G_0 is $V_o = 0.75$ V. The current supplied to the bases of each of the transistors T_1, T_2, \dots, T_N is supplied from the source V_{cc} through the resistor R_c . In this discussion we are assuming that R_c is small enough for adequate current to be available to drive all the transistors T_1, T_2, \dots, T_N to saturation. That is, the current in R_c is equal to

$$\frac{V_{cc} - V_o}{R_c} = \frac{V_{cc} - 0.75}{R_c}$$

Hence, each base current in T_1, \dots, T_N is $I_b = (V_{cc} - 0.75)/NR_c$ if T_1, \dots, T_N are identical. We are assuming that the current is adequate to saturate T_1, \dots, T_N . Note that the collectors of gate G_0 rise no higher than $V_o = 0.75$ V. With at least one input of G_{11} at the logical level 1, the logic level at the output of G_{11} is 0, since $V_{ce} = V_{ce}(\text{sat})$. This voltage level, $V_{ce}(\text{sat})$, is typically between 0.1 and 0.2 V, depending on the current flowing in resistor R_c .

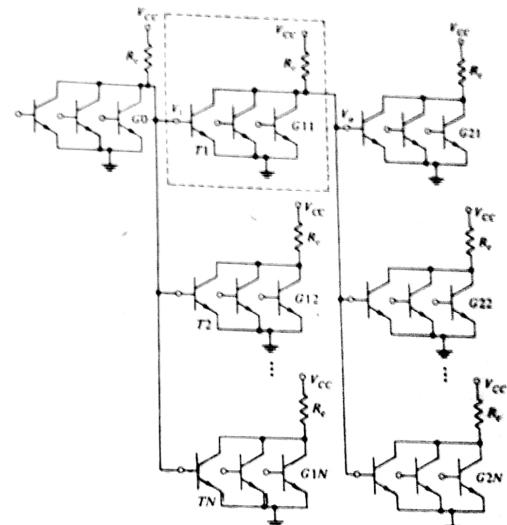


FIGURE 4.2-1
DCTL NOR gate with a fan-out of N .

of the gate. Since, as we have seen, the cut-in point of a silicon transistor occurs at a base-emitter voltage of about 0.65 V, the saturation voltage $V_{ce}(\text{sat}) = 0.1$ or 0.2 V, applied to an input of a succeeding gate is low enough to keep the corresponding transistor cut off.

It is to be observed that the total voltage separation, referred to as the logic swing between the voltages corresponding to the logic levels 1 and 0, is of the order of $0.75 - 0.1 = 0.65$ V.

Input-output characteristic of DCTL The above results are neatly summarized by the (idealized) input-output characteristic shown as the solid plot in Fig. 4.2-2. This figure plots the output V_o of gate G_{11} as a function of an input voltage V_i applied to one of the transistors in the gate. The other remaining transistors in the gate are assumed to be cut off; i.e., the other inputs are in the low state.

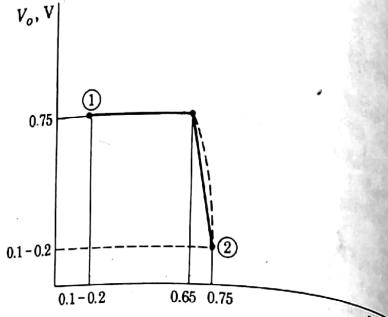


FIGURE 4.2-2
Input-output characteristic of a DCTL NOR gate

At point 1, V_i is at its minimum voltage of $V_{CE}(\text{sat}) = 0.1$ to 0.2 V, depending on the saturation voltage of gate G0. At this input voltage, V_i is equal to 0.75 V due to the loading of gates G21, G22, etc. As V_i increases, T_1 remains cut off until V_i reaches the cut-in voltage of the transistor, which is 0.65 V at room temperature. As V_i increases from 0.65 to 0.75 V, transistor T_1 goes from cutoff to saturation, as shown. V_i is clamped at 0.75 V, and hence the plot is terminated at this voltage (point 2).

A more realistic input-output characteristic is shown by the dashed plot in Fig. 4.2-2. As we have noted (see Sec. 1.7), in the active region the collector current of a transistor I_C is nearly equal to the emitter current I_E , and the emitter current is related to the emitter-base voltage by the diode equation [Eq. (1.7-3)]. Hence, since $V_{BE} = V_i$, we have

$$I_C \approx I_E \approx I_{E0} e^{V_i/V_T} \quad (4.2-1)$$

For values of V_o less than 0.65 the base currents of the transistors being driven by G11 fall to zero and we may write

$$V_o = V_{CC} - R_c I_C = V_{CC} - R_c I_{E0} e^{V_i/V_T} \quad (4.2-2)$$

Hence the input-output characteristic is a falling exponential with ever increasing magnitude of slope. In the range $0.65 \leq V_o \leq 0.75$ V, Eq. (4.2-2) is less exact since it does not take account of the volt-ampere characteristic of the base-collector circuits of the driven transistors.

4.3 CURRENT HOGGING IN DCTL GATES

DCTL gates suffer from a difficulty referred to as *current hogging*. To appreciate this problem we need only recognize that while transistors of similar manufacture are generally quite similar in performance, they are not precisely identical.

Thus, referring to Fig. 4.2-1, suppose that the output of gate G0 is at logic

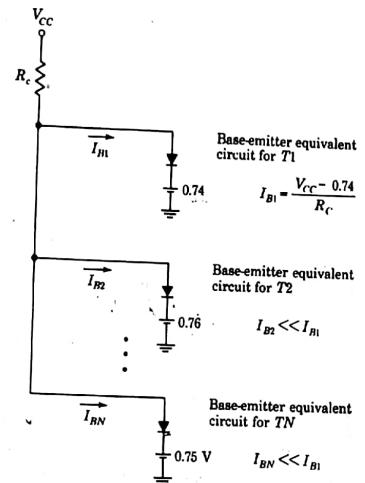


FIGURE 4.3-1
Equivalent circuit to illustrate current hogging in DCTL

level 1, so that T_1 , T_2 , etc., are all to be driven to saturation. Suppose, by way of example, that when T_1 attains saturation, its base-emitter voltage is 0.74 V. But suppose that T_2 requires a base-emitter voltage of 0.76 V for saturation while 0.74 V leaves it in the active region. This difference may be due not only to the fact that the transistors are different but also to the fact that T_1 and T_2 may be in different integrated-circuit packages and hence operating at different temperatures. The temperature difference, in turn, may be due to the different physical location of the two units and possibly a difference in power dissipation in the two integrated circuit packages.

In any event, as seen from Fig. 4.3-1, which shows the output circuit of G_0 and the base-emitter equivalent circuits of T_1 , T_2 , ..., T_N , T_1 will hog much of the available current furnished by V_{CC} through R_c , and other transistors having base-emitter voltages greater than 0.74 V, such as T_2 , will be starved for base current and not be able to attain saturation as required. This current-hogging phenomenon, which is characteristic of DCTL, explains why this type of logic is not presently in wide use, and we shall not consider it further.

4.4 RESISTOR-TRANSISTOR LOGIC (RTL)

The current-hogging difficulty of the DCTL gate can be largely eliminated by introducing resistors in series with the base of each transistor as has been done in the RTL gate of Fig. 4.1-1 (see Prob. 4.4-1). In a medium-power commercial

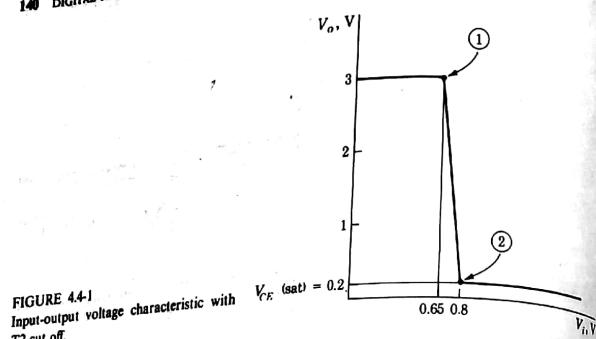


FIGURE 4.4-1
Input-output voltage characteristic with $V_{CE(\text{sat})} = 0.2$.
 T_2 cut off.

gate we find that $R_c = 640 \Omega$ and $R_b = 450 \Omega$, as noted in parentheses in Fig. 4.1-1. The supply voltage is typically $V_{CC} = 3$ V.

The input-output voltage characteristic of the RTL gate is shown in Fig. 4.4-1. Here we have plotted the output V_o as a function of one input V_i under the condition that all other input signals are at voltages which ensure that the corresponding transistor is cut off. Referring to Fig. 4.4-1, we note that if V_i is less than the cut-in voltage, which is 0.65 V at room temperature, T_1 is cut off and $V_o = V_{CC} = 3$ V.

When V_i exceeds 0.65 V, T_1 enters the active region and V_o decreases until T_1 saturates. It is useful to calculate the minimum value of V_i needed to bring T_1 into saturation. For the purpose of illustration let us assume that $V_{CE(\text{sat})} = 0.2$ V, $h_{FE} = 50$, and $h_{FC} = 0.1$. Then from Fig. 1.10-1 we find that $\sigma \equiv I_C/50I_B = 0.85$. Using $R_c = 640 \Omega$, the collector current of T_1 is found from Fig. 4.1-1 to be

$$I_C = \frac{3 - 0.2}{640} = 4.4 \text{ mA} \quad (4.4-1)$$

and hence $I_B = \frac{4.4 \times 10^{-3}}{50(0.85)} \approx 0.1 \text{ mA}$ (4.4-2)

Knowing I_B and assuming that $V_{BE}(T_1) = V_o = 0.75$ V since T_1 is saturated, we can calculate V_i . Referring to Fig. 4.1-1, we have

$$V_i = 450I_B + V_{BE} = 450(0.1 \times 10^{-3}) + 0.75 \approx 0.8 \text{ V} \quad (4.4-3)$$

Further increases in V_i produce only a small change in $V_o = V_{CE}$ since in saturation V_{CE} is typically between 0.1 and 0.2 V. Thus, as seen in Fig. 4.4-1, V_o has dropped to 0.2 V at $V_i = 0.8$ V, and we have idealized the plot by ignoring any further small change in V_o for $V_i > 0.8$ V. The plot is further idealized in that the plot from point 1 to point 2, which should be exponential in character, is drawn as a straight line.

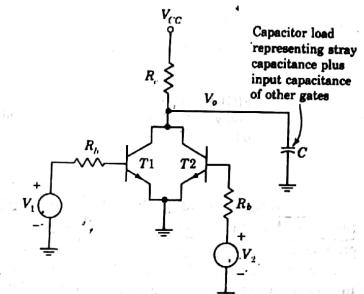


FIGURE 4.4-2
A two-input RTL gate with a capacitive load.

Types of RTL NOR gates As noted, commercially available integrated-circuit RTL NOR gates typically have the collector resistor $R_c = 640 \Omega$ while the base resistors $R_b = 450 \Omega$. A low-power RTL gate is also available having $R_c = 3.6 \text{ k}\Omega$ while $R_b = 1.5 \text{ k}\Omega$. A disadvantage of this low-power gate is that the larger resistors result in slower operation; i.e., the propagation delay time is longer. The reason for this longer delay is that all stray capacitors and capacitors inherent in the active devices must charge and discharge through these larger resistors.

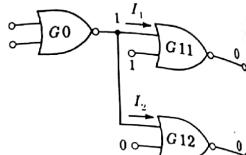
✓ Pull-up resistors The collector resistor R_c in the RTL gate is often called a passive pull-up resistor. The reason for this terminology is to be seen in connection with Fig. 4.4-2, where an RTL gate drives a capacitive load C . This capacitance is due in part to stray capacitance and in part to the capacitance associated with the base-emitter junctions of gates driven by the gate shown.

Assume, initially that V_1 is high and V_2 low, so that T_1 is saturated, T_2 cut off, and V_o in the 0 state. Now let V_1 fall to the 0 state. T_1 now cuts off, and V_o rises toward V_{CC} with a time constant $\tau = R_c C$. We say that V_o is pulled up to V_{CC} by the pull-up resistor R_c . Since R_c is passive, we say that the gate shown employs passive pull-up. Active pull-up is also available for RTL and many of the other gates to be discussed subsequently. The operation of active pull-up is postponed until Sec. 4.7, where we discuss the RTL buffer.

4.5 FAN-OUT

It will generally be necessary, in a physical switching system, for one NOR gate to provide the input logic levels to a number of other such gates. The number of gates driven by a single gate is referred to as the fan-out of that driving gate.

FIGURE 4.5-1
NOR gate G0 driving NOR gates G11 and G12.



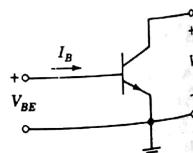
Fan-out is limited by the fact that when the output of the driving gate is at the logic 1 level, the transistors of the driven gate must all be furnished with enough current to saturate them. The current required is supplied by the supply voltage V_{CC} through the collector resistor R_C of the driving gate and hence is limited by the simple constraints of Ohm's law.

It is of some interest to note that when a driving gate, with its limited available driving current, is fanned out to a number of gates, those driven gates which make the greatest drain on the available current are precisely the gates which do not make effective use of the current. This point is made clearer by considering the situation indicated in Fig. 4.5-1. Here a driving gate G0 has its output at logic level 1 and is driving two gates G11 and G12. Gate G0 will provide a current I_1 to the base of a transistor in G11 and a current I_2 to the base of a transistor in G12. In gate G11 the other input is at logic level 1; hence the output of G11 would be logic level 0 even if it were not furnished with current I_1 . In this sense the current I_1 is "wasted." On the other hand, the second input of gate G12 is at logic level 0, and hence the current I_2 is essential to keep the output of G12 at level 0. It is therefore somewhat disconcerting to note that the "wasted" current I_1 is larger than the current I_2 .

To explore this point we need to consider the volt-ampere characteristic seen looking into the base of a transistor operating in the common-emitter configuration. This volt-ampere characteristic is a plot of I_B , the base current, as a function of the base-emitter voltage V_{BE} . This current and voltage are indicated in Fig. 4.5-2. From the Ebers-Moll equations (1.7-3) and (1.7-4) we find that if the transistor is in the active region, the base current $I_B = I_E - I_C$ is related to V_{BE} approximately by

$$I_B = \frac{I_{E0}(1 - \alpha_N)}{1 - \alpha_N \alpha_I} e^{V_{BE}/V_T} \quad (4.5-1)$$

FIGURE 4.5-2
The current I_B and voltage V_{BE} defined.



In arriving at this equation we have taken account of the fact that $e^{V_{BE}/V_T} \gg 1$ and $e^{V_{BE}/V_T} \ll 1$. Since α_I is small ($\alpha_I \approx 0.1$ or even less), we may reasonably carry the approximation somewhat further and write

$$I_B = I_{E0}(1 - \alpha_N) e^{V_{BE}/V_T} \quad (4.5-2)$$

Finally, since $1 - \alpha_N = 1/(h_{FE} + 1) \approx 1/h_{FE}$, we have in the active region

$$I_B = \frac{I_{E0}}{h_{FE}} e^{V_{BE}/V_T} \quad (4.5-3)$$

Next, let us consider that the transistor has been driven to saturation to the maximum extent possible with $\sigma \equiv I_C/h_{FE} I_M = 0$. We then verify, using Eqs. (1.7-3) and (1.7-4), that

$$I_B = I_{E0} e^{V_{BE}/V_T} \quad (4.5-4)$$

Equations (4.5-3) and (4.5-4) appear eminently reasonable. When the transistor is driven to the point where $\sigma = 0$, the collector current may be ignored in comparison with the base current. If we ignore the collector current, we may equivalently consider that the collector terminal of the transistor is floating unconnected. In this case looking into the transistor between base and emitter, we see a simple diode, i.e., the base-emitter junction. The volt-ampere characteristic is then given straightforwardly by the diode equation (4.5-4). However, in the active region, the fraction $h_{FE}/(1 + h_{FE})$ of the emitter-junction current continues on to the collector junction, while the remaining fraction $1/(1 + h_{FE}) \approx 1/h_{FE}$ becomes base current. The base current I_B is again of the form of the diode equation except smaller by the factor h_{FE} , as given by Eq. (4.5-3).

We have referred to the cut-in voltage V_T of a diode as the voltage at which the diode current just becomes large enough to be of significance. It is apparent that the cut-in point of the diode described by Eq. (4.5-3) is higher than the cut-in point of the diode described by Eq. (4.5-4), since in the first case the base current is always smaller by the factor h_{FE} . Since

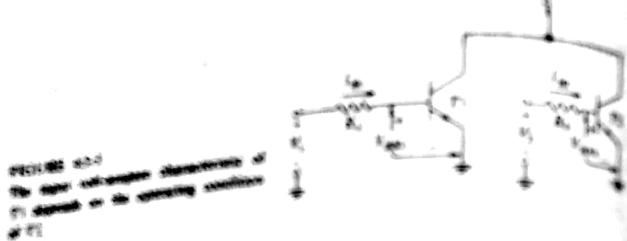
$$\frac{1}{h_{FE}} = \exp\left(\ln \frac{1}{h_{FE}}\right) = \exp(-\ln h_{FE}) \quad (4.5-5)$$

we may write Eq. (4.5-3) in the form

$$I_B = I_{E0} \exp\left[\frac{V_{BE} - V_T \ln h_{FE}}{V_T}\right] \quad (4.5-6)$$

A plot of Eq. (4.5-6) would be identical to a plot of Eq. (4.5-4) except that the plot for Eq. (4.5-6) would be shifted in the positive V_{BE} direction by amount $V_T \ln h_{FE} = 0.1$ V for $V_T = 25$ mV and $h_{FE} = 55$. Thus, in a typical case the cut-in point of the active-transistor base input circuit will be about 0.1 V higher than for a transistor driven to the limit of saturation.

Now consider the situation represented in Fig. 4.5-3, where two transistors are paralleled like a two-input NOR gate. Suppose that T_2 is cut off. Then



The voltage-current characteristic at the base of T_1 is given by Eq. (4.3-3) the presence of T_2 having no effect. But suppose, on the other hand, that T_1 is in saturation with $V_{ce1} = 0.1$ V. In this case when I_1 rises to bring T_1 out of cutoff ($V_{be1} = 0.65$ V), the collector junction will then start forward-biased. That is, with $V_{be1} = 0.65$ V and, say, $V_{ce} = 0.2$, $I_{c1} = 0.65 - 0.2 = +0.45$ V. Hence, T_1 is either cut off or in saturation and is never in the active region. Hence, when T_1 is on, the input voltage-current characteristic at the base of T_1 is given by Eq. (4.3-4).

Now, let us consider the worst-case situation represented in Fig. 4.5-4. Here gate 00 drives one input of N different gates. In every case except one, the transistor paralleling the driver transistor is in saturation. The transistors $T_{12}, T_{13}, \dots, T_{1N}$ are in saturation if they are not cut off. The transistor T_1 is not operating in the active region. Allowing for this worst-case condition, we may estimate the allowable fan-out.

As we have seen, the base current I_{B1} required to drive T_{11} is given by Eq. (4.4-2) as $100 \mu\text{A}$. We can make a rough estimate of the current I_{Bj} by considering that the base voltages of $T_{12}, T_{13}, \dots, T_{1N}$ are less by 0.1 V than the base voltage of T_{11} . Since V_b is common to all transistors and current $I_{Bj} < I_{B1}$ is larger than I_{B1} by $0.1/R_s = 0.1/450 = 220 \mu\text{A}$. Therefore $I_{Bj} = 100 + 220 = 320 \mu\text{A}$

$$I_{Bj} = 100 + 220 = 320 \mu\text{A} \quad (4.5-7)$$

We have also calculated, as given by Eq. (4.4-3), that at room temperature V_b [referred to as V_b in Eq. (4.4-3)] is 0.8 V. The maximum available current from the driver is I_d , given by

$$I_d = \frac{V_{dd} - V_b}{R_s} = \frac{3 - 0.8}{640} = 3.4 \text{ mA} \quad (4.5-8)$$

If the fan-out is to be N , we require that

$$I_d = I_{B1} + I_{B2} + \dots + I_{BN} \quad (4.5-9)$$

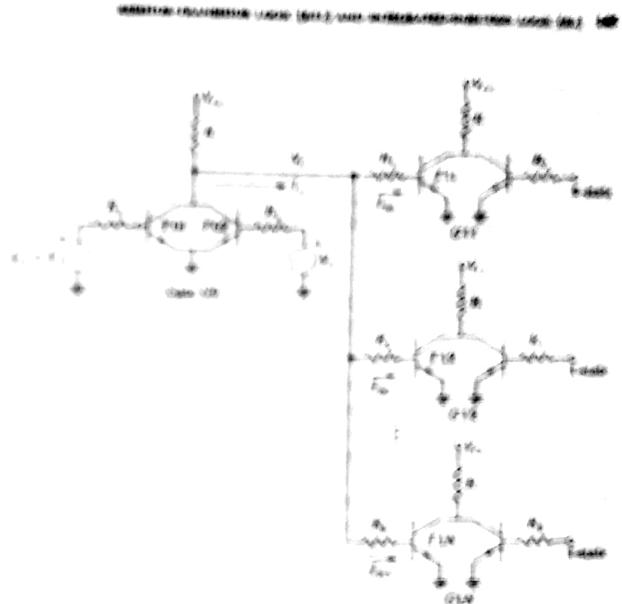


FIG. 4.5-4
Gate 00 drives N two-input gates. In all cases except one the transistor paralleling the driver gate is in saturation.

$$\text{and} \quad 1,400 = 100 + (N - 1)320 \quad (4.5-10)$$

Thus we find that at room temperature

$$N = 11 \quad (4.5-10)$$

At lower temperatures the allowable fan-out would be reduced. This situation results from the temperature sensitivity of the base-emitter voltage $V_{be} = -2 \text{ mV}^\circ\text{C}$) and because k_{TQ} decreases with decreasing temperature. At -55°C we would find that k_{TQ} has fallen to about half its value at room temperature. With k_{TQ} reduced to so low a value, it is easy to verify that at -55°C the fan-out must be reduced to about $N = 7$.

A further possibility may make it necessary to restrict the fan-out even slightly more than indicated by the preceding discussion. Referring to Fig. 4.5-4, we have assumed that when V_b is at logic 1, transistors T_{01} and T_{02} , being cut off, are drawing no current. As a matter of practice it turns out that when these transistors are cut off, each may, in a worst case, draw as much as $75 \mu\text{A}$ each of leakage current (see Sec. 4.9). Suppose then, for example, that gate 00 is a

three-input gate. In this case the leakage current would be $3(75) = 225 \mu\text{A}$. The unavailability of this leakage current as drive current for succeeding gates may well reduce the fan-out by one unit.

Altogether, on the basis of the preceding discussion and taking into account also the variability of transistor and resistor parameters to be anticipated from sample to sample of an integrated circuit, it is not surprising that manufacturers, with befitting conservatism, specify the fan-out of RTL gates at $N = 5$.

4.6 INPUT-OUTPUT VOLTAGE CHARACTERISTIC OF CASCADeD RTL GATES

In connection with logic gates it is often useful to have available a plot of the gate output voltage V_o as a function of the gate input voltage V_i under the circumstances that the gate is fanning out into a number of loads. Such transfer characteristics incorporate in a single figure a great deal of information about operating points, logic swing (the change in voltage levels corresponding to the two logic levels 1 and 0), and compatibility between gate input and output voltages. Such characteristics may also be used to determine a measure of the sensitivity of a gate to noise at any of its inputs and to display the effect of temperature and supply-voltage variation.

We shall plot the transfer characteristic for the RTL gate for a fan-out of 5. That is, referring to Fig. 4.5-4, we shall plot V_o as a function of V_i assuming $N = 5$ and that $T02$ is cut off. We shall also assume that in each of the driven gates all transistors, other than the transistors connected to $G0$, are cut off. Note that in the present situation the base of each right-hand transistor $G1I-G1N$ is to be in the 0-state and not, as appears in the figure, in the 1-state.

When the voltage $V_i = 0 \text{ V}$, the output V_o of gate $G0$ is in the 1 state. All the driven gates, to which the driving-gate fans out, are in saturation. As we have discussed in Sec. 1.9, we shall assume that looking into the base, i.e., between base and emitter, a saturated transistor appears as a voltage source $V_o = 0.75 \text{ V}$ (at room temperature) or more generally [see Eq. (1.2-1)] as a source

$$V_o = 0.75 - (2 \times 10^{-3})(T - 25^\circ\text{C}) \quad (4.6-1)$$

The equivalent circuit needed to calculate the collector voltage V_o of gate $G0$ is as given in Fig. 4.6-1. In this circuit we have implicitly assumed that all driven transistors are identical. Applying Kirchhoff's laws to the circuit of Fig. 4.6-1, we find that (since $450/5 = 90 \Omega$)

$$V_o = \frac{3}{640 + 90} 90 + \frac{V_o}{640 + 90} 640 \quad (4.6-2)$$

Transistor specifications are usually presented at three temperatures, -55°C , room temperature $T = 25^\circ\text{C}$, and $+125^\circ\text{C}$. The low and high temperatures represent the extreme operating temperatures, while room temperature represents

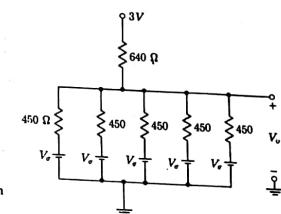


FIGURE 4.6-1
Equivalent circuit to calculate V_o when
 $G0$ is OFF.

typical operating conditions. The output voltage V_o at these three temperatures is found, using Eqs. (4.6-1) and (4.6-2), to be

$$V_o = \begin{cases} 0.85 \text{ V} & T = 125^\circ\text{C} \\ 1.03 \text{ V} & T = 25^\circ\text{C} \\ 1.17 \text{ V} & T = -55^\circ\text{C} \end{cases} \quad (4.6-3)$$

As V_i is increased from 0 V , a point will be reached where transistor $T01$ enters the active region. Again, as discussed in Sec. 1.2, we shall consider that this cut-in point occurs at

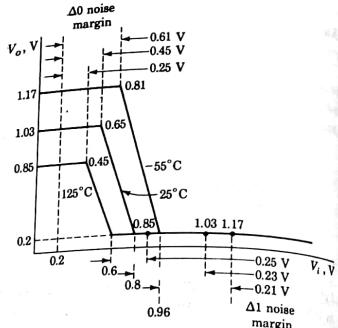
$$\begin{aligned} V_{BE} &= V_o = 0.65 - (2 \times 10^{-3})(T - 25^\circ) \\ &= \begin{cases} 0.45 & T = 125^\circ\text{C} \\ 0.65 & T = 25^\circ\text{C} \\ 0.81 & T = -55^\circ\text{C} \end{cases} \end{aligned} \quad (4.6-4)$$

Increasing V_i further, we find that V_o decreases and eventually will drop to $V_{CE(\text{sat})}$, which we shall again assume to be in the range 0.2 to 0.1 V . The base-emitter voltage of the driving-gate transistors will again be $V_{BE} = V_o$ at saturation.

The input voltage at the point where $T01$ in gate $G0$ reaches saturation has already been calculated in Eq. (4.4-3) to be $V_i = 0.8 \text{ V}$ at room temperature. Using Eq. (4.6-1), we can similarly find V_i at other temperatures:

$$V_i = \begin{cases} 0.6 \text{ V} & T = 125^\circ\text{C} \\ 0.8 \text{ V} & T = 25^\circ\text{C} \\ 0.96 \text{ V} & T = -55^\circ\text{C} \end{cases} \quad (4.6-5)$$

Using the results given in Eqs. (4.6-3) to (4.6-5), we can construct the plots shown in Fig. 4.6-2. To see how these plots are obtained consider $T = 25^\circ\text{C}$. Then for $V_i \leq 0.65 \text{ V}$ (cut-in), $V_o = 1.03$ [Eq. (4.6-3)]. The value of cut-in is given in Eq. (4.6-4). The value of V_i needed to have $V_o = 0.2 \text{ V}$ (saturation) is $V_i = 0.8$, which is obtained from Eq. (4.6-5). The maximum value of V_i is 1.03 V

FIGURE 4.6-2
Input-output characteristic.

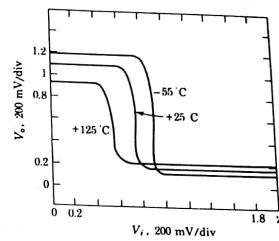
and is obtained by assuming that gate G_0 is driven from another gate (not shown) having the same values of V_o that G_0 has. Using a similar argument, we find that V_i cannot be less than 0.2 V, which occurs when the gate driving G_0 saturates.

These plots are, of course, idealized and representative rather than real and exact. However, they are close enough to the actual plots that would be obtained in a representative real situation to be of great value. By way of comparison we compare the plots of Fig. 4.6-2 with the plots of Fig. 4.6-3 determined experimentally for typical gates. The real plots indicate some dependence of the temperature on the transistor saturation collector-emitter voltage. This, of course, should be expected, since, referring to Eq. (1.10-3), we see that $V_{CE}(\text{sat})$ is proportional to $V_T = kT/q$ and therefore $V_{CE}(\text{sat})$ increases with an increase in temperature. Our idealized plots have ignored this temperature dependence since the variation is small. We also note that while we have assumed that $V_o(\text{sat}) = 0.2$ V at room temperature, the characteristic shown in Fig. 4.6-3 has a value of 0.16 V. Starting with $V_{CE}(\text{sat}) = 0.16$ V at room temperature and since $0^\circ\text{C} = 273^\circ\text{K}$, we then have

$$V_{CE}(\text{sat}, 125^\circ\text{C}) = \left[\frac{273 + 125}{273 + 25} \right] 0.16 = 0.21 \text{ V} \quad (4.6-6)$$

and we correspondingly find $V_{CE}(\text{sat}, -55^\circ\text{C}) = 0.12$ V. These results agree with Fig. 4.6-3.

Noise margin. The plots of Fig. 4.6-2 are useful because they provide the input and output voltage ranges corresponding to the two logic levels 1 and 0. We keep in mind that generally the input of one gate is the output of a preceding gate. Thus, when the logic level 0 appears at the output of a driving gate and

FIGURE 4.6-3
Typical RTL-gate transfer characteristic for three temperatures at fan-out = 5.

hence at the input to a driven gate, this logic level will correspond to the voltage $V_{CE}(\text{sat}) \approx 0.2$ V. We now note from Fig. 4.6-2 that such a voltage, 0.2 V, is low enough to represent properly the logic level 0. For over the entire range of temperature contemplated in Fig. 4.6-2, the driven gate transistor will be cut off and the output of the driven gate will therefore be at a high voltage. Further, we note that this voltage $V_{CE}(\text{sat}) = 0.2$ V provides a margin of safety in assuring cutoff. For even in the least assured case, which occurs at $T = 125^\circ\text{C}$, there is a voltage difference $0.45 - 0.20 = 0.25$ V between $V_i = V_{CE}(\text{sat})$ and the minimum input voltages which would allow the driven transistor to enter its active region, $V_i = 0.45$ V. Having marked off the voltage $V_{CE}(\text{sat}) = 0.2$ V on the abscissa of Fig. 4.6-2, we can indicate the margins of safety for the three temperatures in the manner shown in the figure. This margin of safety is also referred to as the noise margin for an input corresponding to logic level 0 and is represented by $\Delta 0$. We have that

$$\Delta 0 = 0.65 - k(T - 25^\circ\text{C}) - V_{CE}(\text{sat}) = 0.45 - k(T - 25^\circ\text{C}) \quad (4.6-7)$$

where k , as noted, is $k = 2 \text{ mV}/^\circ\text{C}$. We note the decrease of $\Delta 0$ with increasing temperature.

The sensitivity to temperature of $\Delta 0$ is actually somewhat larger than indicated in Eq. (4.6-7), for we see in Fig. 4.6-3 that $V_o = V_{CE}(\text{sat})$ increases somewhat with temperature. Thus, the single dashed line erected vertically from the abscissa at $V_i = V_{CE}(\text{sat}) = 0.2$ V in Fig. 4.6-2 should be replaced by three lines, corresponding to the three temperatures. The reduction in $\Delta 0$ below the values given in Fig. 4.6-2 would be greatest for $\Delta 0$ at 125°C , the case where there is already the smallest noise margin. We may, as a matter of fact, observe that this limited $\Delta 0$ encountered with increasing temperature constitutes a disadvantage of consequence in RTL circuitry.

Now let us turn our attention to the case in Fig. 4.5-4, where V_i has increased sufficiently for $T01$ to be driven into saturation. In this case, of course, the higher the voltage V_i the more assurance we have of saturating transistor $T01$.

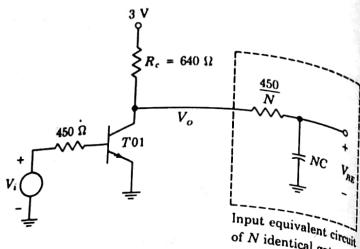


FIGURE 4.6-4
Equivalent circuit to determine the rise time of V_o .

We note from Sec. 4.5 that the fan-out of the gate driving G_0 limits the extent to which V_i can rise, just as the fan-out of G_0 limits the maximum value of V_o . The greater the fan-out the lower the available voltage. Hence, we must consider that the maximum available input voltage to drive a gate is not the unloaded output voltage of a driving gate but the output voltage which is loaded by a maximum fan-out. We have drawn Fig. 4.6-2 for a fan-out of 5 precisely because this fan-out is nominally the maximum intended with RTL gates. The output voltages (1.17, 1.03, and 0.85 V) for the fan-out of 5 are, of course, the input voltages available to drive succeeding transistors to saturation.

We can now mark off on the abscissa of Fig. 4.6-2 these output voltages corresponding to a fan-out of 5. The noise margins $\Delta 1$ for an input corresponding to logic level 1 can be read off as indicated in the figure. Thus, at $T = -55^\circ$ the output is 1.17 V, and the minimum output which will keep a driven transistor in saturation is 0.96 V. The noise margin is therefore $\Delta 1 = 1.17 - 0.96 = 0.21$ V. We note from Fig. 4.6-2, that the $\Delta 1$ noise margins are not sensitively dependent on temperature.

Rise time The rise time of gate G_0 is affected by the number of gates it drives. To illustrate this consider that in Fig. 4.5-4 V_i is in the 1 state and V_o is in the 0 state. Then T_{11} to T_{1N} are cut off, and the base-emitter junction of each of these transistors appears to be a capacitor. Let us approximate this capacitor as a real constant capacitor C . Then the equivalent circuit of G_0 and the driven gates in the region where V_i changes from its 1 state to its 0 state (with V_i remaining in the 0 state) is shown in Fig. 4.6-4. Here we have assumed that each of the $N = 5$ driven gates is identical.

Now let V_i fall instantaneously from its 1 state to its 0 state. T_{01} cuts off, and the base-emitter voltage of T_{11} to T_{1N} rises with a time constant

$$\tau = \left(R_c + \frac{450}{N} \right) NC = (640N + 450)C \quad (4.6-8)$$

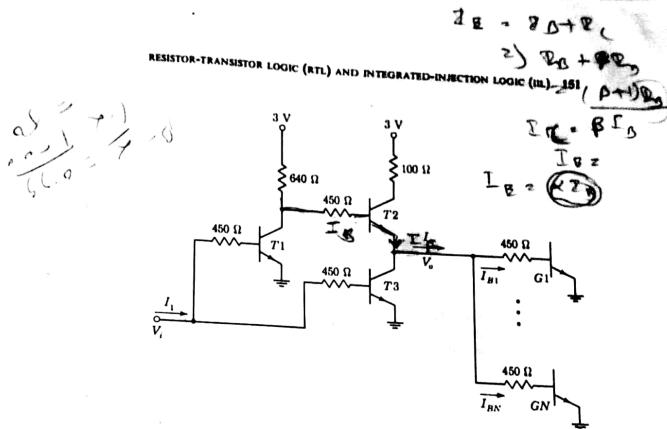


FIGURE 4.7-1
An RTL buffer inverter driving N gates.

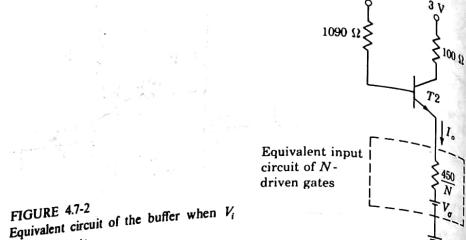
The manufacturer generally specifies that $C \approx 5 \text{ pF}$; thus with $N = 5$, $\tau = 18 \text{ ns}$. In the next section we shall see that an RTL gate using an active pull-up can have the same value for τ and a significantly larger fan-out.

It should be kept in mind that the input capacitance to a transistor is not constant and is a function of the current in the base. Thus, the time constant indicated here is of only qualitative rather than quantitative value.

4.7 AN RTL BUFFER

The RTL gates discussed in the previous sections employed a passive pull-up R_c . This limited the output current, available to drive other gates, to a value less than V_{cc}/R_c . The buffer is an RTL gate using an active pull-up to achieve a very low output impedance. Thus, the output-current capability of the buffer is significantly greater than that of the ordinary gate. As a result, while the ordinary RTL gate has a fan-out of 5, the buffer has a fan-out of 25. In addition, since the buffer has a low output impedance, the rise time when driving a capacitive load is significantly less than that obtained when using a passive pull-up.

A typical RTL buffer is shown in Fig. 4.7-1. The transistor T_2 is the active pull-up which replaces the passive pull-up resistor R_c in the ordinary logic gate. The transistor T_1 serves as a logic inverter. When V_i is at logic 0, the output of T_1 is at logic 1 and vice versa. Thus, the logic levels at the bases of T_2 and T_3 are always different, and when one of these transistors is conducting, the other is cut off. When T_2 is OFF, its collector current is zero; T_3 is saturated

FIGURE 4.7-2
Equivalent circuit of the buffer when V_i is in the 0 state.

in the extreme with $\sigma = I_C/h_{FE} I_B = 0$, and (as can be seen in Fig. 1.10-1) its voltage $V_{CE}(\text{sat}) = 60 \text{ mV}$.

When T_3 is off and T_2 is on and driving N gates, its output current I_o can be calculated from the circuit of Fig. 4.7-2. Assume first that T_2 is in its active region with $V_{BE}(T_2) = 0.7 \text{ V}$ and assume also $h_{FE} = 50$. At room temperature, as usual, we take $V_o = 0.75 \text{ V}$. Applying Kirchhoff's voltage law to the loop which includes the base-emitter junction of T_2 , we find

$$I_o = \frac{3 - 0.7 - 0.75}{1090/50 + 450/N} = \frac{7.1 \times 10^{-2}}{1 + 20.6/N} \quad (4.7-1)$$

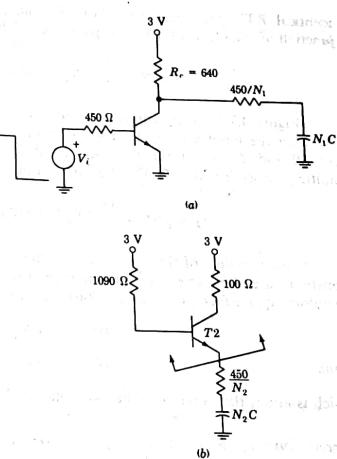
If, instead, we assume that T_2 is saturated, we have $V_{BE} = 0.75 \text{ V}$ and $V_{CE} = 0.2 \text{ V}$. In this case we find (Prob. 4.7-4) that

$$I_o(\text{sat}) = \frac{21.8 \times 10^{-3}}{1 + 4.9/N} \quad (4.7-2)$$

To tell whether or not T_2 is saturated we need only compare the two currents. For if $I_o(\text{sat}) > I_o$, the transistor is not saturated, while if $I_o(\text{sat}) < I_o$, the transistor is saturated. It can be verified that the transistor is not saturated for $N = 1$ and is saturated for $N \geq 2$.

The buffer is ordinarily employed to drive a large number of gates. We shall therefore assume that $N \geq 2$ and that the transistor T_2 is saturated. Since $I_o(\text{sat})$ is supplied to all N gates, the current supplied to each gate is $I_{g1} = I_{g2} = I_{gN} = I_B$, where

$$I_B = \frac{I_o(\text{sat})}{N} = \frac{21.8 \times 10^{-3}}{N + 4.9} \quad (4.7-3)$$

FIGURE 4.7-3
(a) RTL gate circuit to determine time constant τ_t . (b) Equivalent circuit of buffer just after T_1 and T_3 cut off.

For the purpose of comparing the fan-out of the buffer to the fan-out of the ordinary RTL gate we assume that $I_B = 320 \mu\text{A}$, as in Eq. (4.5-7). Then

$$\frac{21.8 \times 10^{-3}}{N + 4.9} \geq 0.32 \times 10^{-3} \quad (4.7-4a)$$

and

$$N \leq 65 \quad (4.7-4b)$$

Note that in Eq. 4.5-10 we found $N = 11$. Thus, we have improved the fan-out by a factor of 6. Manufacturer's specifications, typically conservative, state that the maximum fan-out of a buffer is $N = 25$, which is a factor of 5 greater than the specification of fan-out for an RTL gate.

We compare now the rise times of a buffer with a passive pull-up NOR gate. The rise time of a gate is proportional to the RC time constant of the gate and load. In Fig. 4.7-3a we have the equivalent circuit of an RTL gate driving N_1

identical RTL gates (see Sec. 4.6). We have represented the base-emitter junction of each transistor by a capacitor C . The time constant τ_1 is

$$\tau_1 = \left(640 + \frac{450}{N_1}\right) N_1 C = (640N_1 + 450)C \quad (4.7-5)$$

Figure 4.7-3b is the equivalent circuit of a buffer, driving N_2 RTL gates, just after the input voltage V_i dropped from its 1 to 0 state. T_1 and T_3 are cut off and T_2 is saturated. Since T_2 is saturated, looking back into the emitter, we see just the 100- Ω collector resistor. Hence, the time constant is

$$\tau_2 \approx \left(100 + \frac{450}{N_2}\right) N_2 C = (100N_2 + 450)C \quad (4.7-6)$$

If the rise time of the RTL gate and buffer are to be the same, their time constants are the same. To obtain a relationship between N_1 and N_2 , we therefore equate Eqs. (4.7-5) and (4.7-6), which yields

$$640N_1 = 100N_2 \quad (4.7-7a)$$

$$N_2 = 6.4N_1 \quad (4.7-7b)$$

Thus which is about the same ratio that we obtained from static considerations.

Current-limiting Referring to Fig. 4.7-2, we note that with T_2 saturated, the collector current is limited by the 100- Ω collector resistor. This resistor is designed to allow T_2 to saturate and yet limit its power dissipation.

To illustrate, let T_2 be saturated so that $V_{CE} \approx 0.2$ V. Then with $N = 25$, $I_c(\text{sat}) = 18$ mA. We can show (see Prob. 4.7-3) that $I_B = 1$ mA, so that $I_C = 17$ mA. Thus, the collector dissipation in T_2 is $P_C = 3.4$ mW. However, if the 100- Ω resistor were short-circuited, T_2 would not saturate, since $V_{CE} > 0$. Then from Eq. (4.7-1), I_C ($N = 25$) = 39 mA, and from Fig. 4.7-2,

$$V_{CE} = 3 - (450/25)(39 \times 10^{-3}) - V_o = 1.55 \text{ V} \quad (4.7-8)$$

The dissipation is now 60 mW, a considerable increase and a sizable dissipation for an integrated-circuit transistor.

4.8 AN RTL EXCLUSIVE-OR GATE

The EXCLUSIVE-OR gate was introduced in Sec. 3.10, where it was defined as a device which performs the operation

$$Z = \overline{A} \cdot B + A \cdot \overline{B} \equiv A \oplus B \quad (4.8-1)$$

An RTL circuit capable of performing this logic operation is shown in Fig. 4.8-1a. The operation of this circuit is as follows. Consider that B is in the

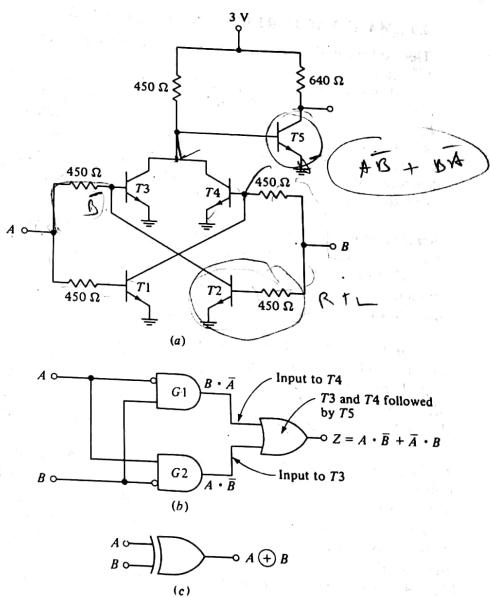


FIGURE 4.8-1
(a) An EXCLUSIVE-OR circuit using RTL.
(b) Logic-gate representation and (c)
symbol for EXCLUSIVE-OR circuit.

logic 0 state. Then T_2 is OFF. Thus, the input to T_3 is A . If, however, B is in the 1 state, T_2 is saturated and T_3 is cut off. Hence, the input to T_3 is $A \cdot \overline{B}$. Similarly, the input to T_4 can be shown to be $\overline{A} \cdot B$. These two operations are represented by G_2 and G_1 , respectively, in Fig. 4.8-1b.

Referring to Fig. 4.8-1a, we see that transistors T_3 and T_4 represent a NOR operation. Transistor T_5 inverts the NOR output, producing an OR gate. This is shown in Fig. 4.8-1b. Thus, the output of T_5 is $Z = A \cdot \overline{B} + \overline{A} \cdot B$, as required.

4.9 MANUFACTURER'S SPECIFICATIONS

This section is devoted to a discussion of the manufacturer's specifications of RTL gates.

There are four basic types of RTL gates. They are the medium-power gate (MRTL), where $R_c = 640 \Omega$ and $R_b = 450 \Omega$; the low-power gate (LRTL), where $R_c = 3.6 \text{ k}\Omega$ and $R_b = 1.5 \text{ k}\Omega$; the buffer, which is available in either the medium- or low-power class; and the EXCLUSIVE-OR gate. Some of the more important characteristics of each of these gates are presented in Fig. 4.9-1 for room temperature, $T = 25^\circ\text{C}$, and for a typical gate. Corresponding specifications at -55 and $+125^\circ\text{C}$ are also available in manufacturer's literature.

Input current, I_{in} . This is the maximum current that will be drawn by a gate input. We have estimated [Eq. (4.5-7)] that this current is about $320 \mu\text{A}$. The manufacturer even more conservatively allows a worst case of $435 \mu\text{A}$. This manufacturer's specification is about right to account for his specification that the allowable fan-out is 5.

Output current, I_{out} . This is the minimum current available to drive other gates under the circumstance of a fan-out of N . We note from Fig. 4.9-1 that for an MRTL gate, $I_{out} = 2.54 \text{ mA}$. Since $I_{in} = 435 \mu\text{A}$, we should expect that $I_{out} = 5I_{in}$. We find that $5I_{in} = 5 \times 435 \mu\text{A} = 218 \text{ mA}$. The difference $2.54 - 2.18 = 0.36 \text{ mA}$ provides an extra margin of safety and allows some current for leakage.

Leakage current. When a transistor input is at logic 0, we would expect the collector current to be zero. However, such is not always the case since the base-emitter voltage is always somewhat positive. For example, if the transistor is driven by a gate which is saturated, then the base-emitter voltage of the transistor is equal to the saturation voltage of the driving gate. In a worst-case situation this saturation voltage may be as high as 0.4 V . As can be seen in Fig. 4.9-1, the manufacturer guarantees that the leakage current I_L is always less than $218 \mu\text{A}$ for the MRTL gate and buffer and $100 \mu\text{A}$ for the LRTL gate and buffer.

Input and output loading factors. To provide guidance concerning allowable fan-out of gates, buffers, etc., manufacturers often assign numbers to the input and output terminals. These numbers, called input and output loading factors, are proportional (to the nearest integer on the conservative side) to the current required by an input or available from an output. Thus, as appears in Fig. 4.9-1, the input loading factor of an MRTL gate is 1, and the output loading factor N_G is 5. These numbers are interpreted to mean that the available output current is at least 5 times (but less than 6 times) the required input current. Consequently, when a gate drives a gate, the allowable fan-out is 5. We note that a buffer has an input loading factor of 2. Hence, a gate may be fanned out to two buffers but not three. The allowable fan-out of an LRTL gate is 4, and hence the input and output loading factors are 1 and 4, respectively.

	MRTL			LRTL	
	Gate	Buffer	Exclusive-or	Gate	Buffer
Fan out (N_G)	5		5	4	
Propagation delay time, t_{pd}	12 ns	20 ns	12 ns	27 ns	57 ns
Power dissipation, P_d					
Inputs high	19 mW	16 mW	72 mW	4.8 mW	5.5 mW
Inputs low	5 mW	45 mW	72 mW	0.5 mW	16 mW
Input current, I_{in}	435 μA	870 μA	870 μA	130 μA	260 μA
Output current, I_{out}	2.54 mA	12.7 mA	2.54 mA	815 μA	4 mA
Output leakage current, I_L	218 μA	218 μA	218 μA	100 μA	100 μA
Saturation voltage	210 mV	210 mV	210 mV	220 mV	220 mV

FIGURE 4.9-1
Manufacturer's specifications.

The loading factors are modified when MRTL and LRTL gate families are used together. The loading factors of the LRTL gate remain 1 and 4, but the loading factors of the MRTL gate become 3 and 16. These numbers indicate that an MRTL gate is able not only to drive five other MRTL gates but can simultaneously drive an LRTL gate, that is, $16 = 5 \times 3 + 1$. In general, in fanning gates out we need only observe the rule that the output loading factor of the driving gate must be equal to or less than the sum of the input loading factors of the driven gates.

4.10 PARALLELING RTL GATES

To increase the number of inputs available in an RTL gate we can operate these gates in parallel. We consider now how the input and the output loading factors are affected by such paralleling.

One way of paralleling is shown in Fig. 4.10-1, where we have connected the output terminals of the two gates A and B and in each case have connected the top end of the collector resistor R_c to the supply voltage V_{cc} . As a consequence, the composite gate operates with a collector resistor $R_c/2$. It will be recalled that the output loading allowed on an RTL gate is limited by the fact that the output current required to drive other gates must be furnished from V_{cc} through the collector resistor. Since the paralleling has reduced the collector

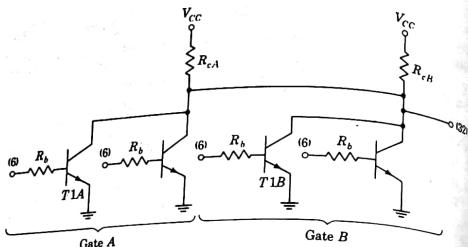


FIGURE 4.10-1
Paralleling two gates on the same chip.

resistor by a factor of 2, the effect of the paralleling is to increase the output loading factor by this same factor of 2. If MRTL gates are involved which individually have an output loading factor of 16, then, as indicated in Fig. 4.10-1, the paralleling has increased the output loading factor to 32. Altogether, then, the paralleling has doubled the output loading factor and has doubled as well the number of available input terminals to the gate.

The price to be paid for these advantages, as noted in Fig. 4.10-1, is that the input loading factor must also be increased by a factor of 2, increasing it to 6, while for a single gate the input loading factor is 3. This increased input load results from the fact that the collector resistance is $R_c/2$ rather than R_c . For with $R_c/2$ the collector saturation current increases by a factor of 2, and the base current required to drive a transistor to saturation must increase correspondingly by 2 (for the same α and the same saturation voltage). Similarly if N gates are paralleled, the output loading factor becomes $16N$ and the input loading factor becomes $3N$.

These results for paralleling apply when the individual gates involved are part of the same integrated circuit, i.e., are on the same chip. When the individual dual gates are on different chips, the input loading factor must be increased beyond $3N$ for a reason now to be discussed.

A characteristic of integrated circuits is that there may be considerable variability in the values of resistors in any given circuit. Thus, from sample to sample, in an RTL gate, R_c and R_b may vary widely. On the other hand, there is substantially less variability in the ratio of resistors R_c/R_b . This greater uniformity of the resistor ratio serves to provide a measure of compensation in the operation of a gate operating individually. For suppose, in a particular gate, that R_c is considerably larger than average. Then in a typical operating situation the base current of a transistor in such a gate will be smaller than average. On the other hand, if R_b is larger, R_c will be also. Hence, the transistor saturation current will be smaller, and the smaller available base current will be

adequate to drive the transistor to saturation. Similarly if R_c is smaller, the saturation current will be larger. But the correspondingly smaller R_b will serve to increase the base current, as required. Finally we may note that if the resistors in a gate depart from the average, all the resistors in all the gates on that same chip depart from the average in the same direction.

Now let us consider the situation when the paralleled gates are on different chips. In this case the resistor on one gate may be high and on the other gate low. Thus, for example, in Fig. 4.10-1, consider that in gate *A* the collector resistor (call it R_{cA}) is high and in gate *B*, R_{cB} is low. Suppose that the paralleled gate is to be driven to logic level 0 by driving transistor *T1A* to saturation. Since R_{cB} is smaller than $R_{cA}/2$, the parallel combination of R_{cA} and R_{cB} is less than $R_{cA}/2$. Therefore, when gate *B* is paralleled with gate *A*, the current required to drive *T1A* to saturation is more than doubled. Hence, the input loading factor at the input to *T1A* is also more than doubled by the paralleling. Thus, the input loading factor must be increased above 6. To put the matter another way, we say that to drive the extra current required in R_b of *T1A* we must assure a higher voltage at the input to *T1A*. This result can be accomplished by restricting somewhat the allowed loading on the driving gate. If we specify a higher input loading factor for gate *A*, we shall have accomplished precisely this end.

Referring again to Fig. 4.10-1 and assuming again that $R_{cB} < R_{cA}$, if the drive is to be applied to the base of transistor *T1B* instead of *T1A*, the loading factor at this input might be set at less than 6. On the other hand, when gates are paralleled, unless we take special pains to investigate, we shall not know which gate has the higher resistors and which the lower. Hence, the only thing we can do is to allow an extra margin of input loading factor at every gate input.

In the matter of the extra margin of input loading factor we must seek guidance from the manufacturer, who will base his recommendations on what he knows his manufacturing tolerances to be. Typically, we find that in MRTL, when a gate on a first chip is to be paralleled with a gate on a second chip, it is recommended that the input loading factor be increased by 0.75 load. Thus, in Fig. 4.10-1 we would set the input loading factors at 6.75. Finally consider that a total of N gates are to be paralleled with N_A of the gates on chip *A* and N_B of the gates on chip *B* ($N_A + N_B = N$). Then, by an easy extension of the present discussion it would appear that the input loading factor on chip *A* is $3N + 0.75N_B$ and on chip *B* is $3N + 0.75N_A$.

4.11 SPECIFICATION OF OPERATING VOLTAGES

Figure 4.6-2 indicated the input-voltage-output-voltage variation of an RTL gate having a maximum fan-out of 5, a supply voltage $V_{cc} = 3$ V, and operating temperatures of -55 , $+25$, and $+125^\circ\text{C}$. However, the figure does not take account of the variability of transistor and resistor parameters due to manufacturing tolerances. To take account of such variability the manufacturer specifies

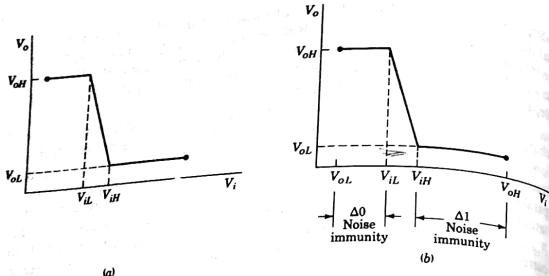


FIGURE 4.11-1
Input-output characteristic showing (a) worst-case parameters and (b) $\Delta 0$ and $\Delta 1$ noise immunities.

the worst-case parameters V_{oH} , V_{iH} , V_{oL} , V_{iL} shown in Fig. 4.11-1a. These parameters are specified at a given temperature, $V_{CC} = 3V \pm 10$ percent, maximum fan-out, and worst-case manufacturing tolerances. The definitions of these parameters are:

V_{oH} The minimum voltage which will be available at a gate output when the output is supposed to be at logic 1

V_{iH} The minimum gate input voltage which will unambiguously be acknowledged by the gate as corresponding to logic 1

V_{oL} The maximum voltage which will appear at a gate output when the output is supposed to be at logic 0

V_{iL} The maximum gate input voltage which will unambiguously be acknowledged by the gate as corresponding to logic 0

Thus, if V_i and V_o are gate input and output voltages, respectively, it is guaranteed that if $V_i \leq V_{iL}$, then $V_o \geq V_{oH}$ and if $V_i \geq V_{iH}$, then $V_o \leq V_{oL}$.

In Fig. 4.11-1b we have marked off the voltages V_{oL} and V_{oH} on the input voltage axis. It now appears that the noise immunities are

$$\Delta 0 = V_{iL} - V_{oL} \quad \text{and} \quad \Delta 1 = V_{oH} - V_{iH}$$

Since these four parameters V_{oH} , V_{iH} , V_{oL} and V_{iL} completely describe the worst-case input-output voltage characteristic and determine the $\Delta 0$ and $\Delta 1$ noise immunity, there is hardly any need for plotting the characteristic of Fig. 4.11-1. Instead, the same information is often presented in the manner shown in Fig. 4.11-2.

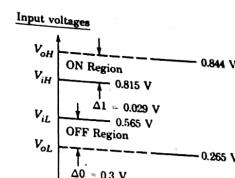


FIGURE 4.11-2
Manufacturer's specified noise margins.

EXAMPLE 4.11-1 For a given RTL gate operating at 25°C, the worst-case parameters as specified by the manufacturer are $V_{oH} = 844$ mV, $V_{iH} = 815$ mV, $V_{iL} = 565$ mV, and $V_{oL} = 265$ mV. Find the $\Delta 0$ and $\Delta 1$ noise immunity.

SOLUTION The worst-case parameter values are shown in Fig. 4.11-2. Thus, the worst-case $\Delta 0$ noise immunity is 300 mV, while the worst-case $\Delta 1$ noise immunity is 29 mV. These results are significantly poorer than those shown in Fig. 4.6-2, where the $\Delta 0$ noise margin is 450 mV and the $\Delta 1$ noise margin is 230 mV. These results seem worse than they really are. The manufacturer readily concedes that the results are extremely conservative and that larger noise margins actually exist.

4.12 PROPAGATION DELAY TIME

We have noted (Sec. 1.17) that there are propagation delays associated with logic gates. In RTL gates these delays are generally specified in terms of the parameters defined in Fig. 4.12-1. Here an input waveform (presumably the output of a preceding gate) is shown which makes a transition from logic 0 to logic 1 and thereafter a reverse transition. The corresponding output waveform is also shown. As is indicated, propagation delays are measured from the points on the waveform which are 0.5 V positive with respect to the logic 0 level. Since logic 0 is at about 0.2 V, the points on the waveform are at 0.7 V, which is very nearly the midpoint of the voltage range at which the transistors of the gates are passing through their active region (see Fig. 4.4-1). The time $t_{pd}(HL)$ is the propagation delay associated with a transition of the output waveform from its higher to its lower voltage while $t_{pd}(LH)$ is the corresponding delay as the output swings from its lower to its higher voltage level. The two propagation delays are not ordinarily equal.

Propagation delays result, in part, from the fact that, as voltages change, capacitors must be charged and discharged. The total capacitance with which the gate must contend depends on the fan-out and also on the fan-in (i.e., the number of inputs on the gate). Hence the propagation times are themselves

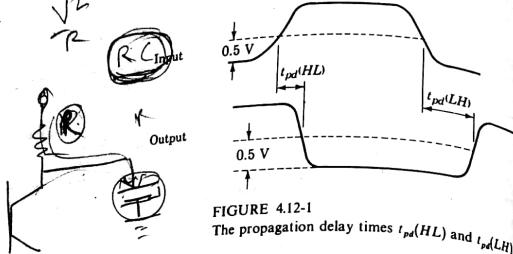


FIGURE 4.12-1
The propagation delay times $t_{pd}(HL)$ and $t_{pd}(LH)$

functions of fan-out and fan-in. Typically in RTL gates propagation times are of the order of 10 ns. Such times compare favorably with propagation times associated with other types of gates yet to be considered. Unfortunately, in spite of the favorable propagation times, RTL gates are not widely used because of their relatively poor noise margins and fan-out capabilities.

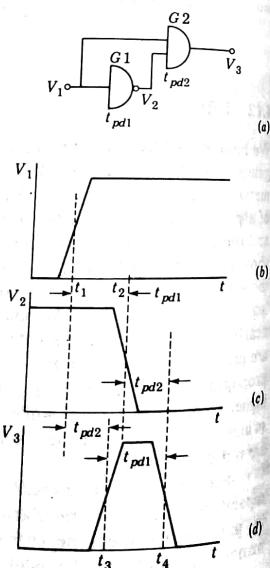


FIGURE 4.12-2
(a) A gate structure whose response is adversely affected by the propagation time delay t_{pd1} . (b) The input waveform. (c) The output of G_1 . (d) The output pulse from G_2 .

Propagation delay hazards As an example of the difficulty which may be caused by propagation delay in gates, consider the situation represented in Fig. 4.12-2. The NAND gate (used as an inverter) has a propagation delay t_{pd1} , and the AND gate a delay t_{pd2} . It is intended that V_3 shall be in the 0 state independently of V_1 , as is indeed consistent with the logic of the circuit. However, as we may see from the waveforms in the figure, a transition in V_1 from logic 0 to logic 1 will result in a positive pulse of duration t_{pd1} as appears in (d). V_1 makes its transition at t_1 , taken for simplicity, to be at the midpoint of the voltage excursion of the waveform V_1 , the waveform V_1 having a finite rise time since it itself is the output of some preceding gate (not shown). V_2 then makes an excursion from 1 to 0 after a time t_{pd1} at time t_2 . (We take the fall time of V_2 to be about the same as the rise time of V_2 .) Since now V_1 rises to 1 before V_2 falls from 1, there will be an interval when the AND gate output V_3 will be 1. V_3 rises to 1 at $t = t_3$ delayed by an interval t_{pd2} from t_1 . At $t = t_2$, after a delay t_{pd1} , V_2 falls back to 0 and finally, at $t = t_4$, after a delay again of t_{pd2} , V_3 will fall back to 0. In summary, it then appears that the step in V_1 has given rise to a pulse in V_3 . Note that the time of occurrence of the pulse is determined by t_{pd2} but that the width of the pulse is determined by t_{pd1} . Furthermore, the very existence of the pulse results from the delay t_{pd1} . Unintended results due to such gate propagation delays are referred to as *hazards*.

Charge compensation to reduce propagation time As noted, propagation delays result, in part, from the necessity to establish and remove base charge as transistors are turned ON and OFF. (See Sec. 1.17.) One way by which this process may be hastened is to provide for the flow of impulsive (very large albeit short-duration) currents into and out of the base. In RTL, such impulsive currents may be provided by bridging capacitors across the base resistors R_b , as shown in Fig. 4.12-3a.

Suppose T_1 is in saturation because its input is at logic level 1. When V_1 drops to logic 0, T_1 should cut off as the stored base charge is reduced to zero. In the absence of the capacitor C , this charge must dissipate in part through recombination in the base and in part by flowing out of the base through R_b . With the capacitor, however, an impulsive current can flow out of the base, removing the base charge much more rapidly. If the capacitor is adequately large it is possible, in principle at least, to transfer all the base charge to the capacitor instantaneously thereby turning the transistor OFF with limitless speed. The use of capacitors to draw charge abruptly out of a semiconductor is called *charge compensation*.

In integrated circuitry, conventional capacitors are used only infrequently because they take up a large amount of area on the IC chip. However, the junction capacitance of a transistor is sometimes employed when capacitors are required. In Fig. 4.12-3b such junction capacitors have been bridged across the resistors R_b through the addition of transistors T_3 and T_4 . The capacitances of the base-emitter junctions of these added transistors bridge the resistors R_b .

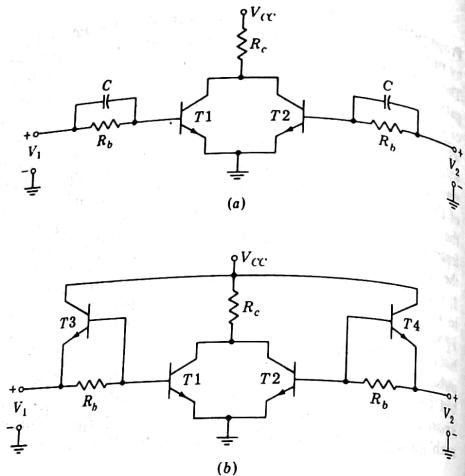


FIGURE 4.12-3
 (a) An RTL gate with capacitors added to provide charge compensation. (b) A practical application of charge compensation using the capacitances across the base-emitter junctions of the added transistors.

A family of RTL gates as in Fig. 4.12-3b with $V_{CC} = 4.0$ V, $R_c = 1.9$ k Ω , and $R_b = 1.2$ k Ω is manufactured by the Western Electric Co. for use by the Bell System.

4.13 INTEGRATED-INJECTION LOGIC (IIL)

We have seen in Sec. 4.5 that when transistors are paralleled, as in RTL gates, the base current drawn by one transistor becomes inordinately large if one of its paralleling transistors goes to saturation. This increase in base current becomes progressively more pronounced as more of the paralleling transistors are turned on and may be particularly pronounced in high fan-in gates. This feature of operation of paralleled transistors accounts, in large measure, for the relatively low fan-out of RTL gates. In addition DCTL suffers from the difficulty of current hogging, discussed in Sec. 4.3. For these reasons DCTL has never found any extensive applications. And while RTL did enjoy a brief period of popularity, it has now fallen into disfavor and is not presently incorporated into new digital systems.

In succeeding chapters we shall explore a number of other families of logic which do not have the limitations of DCTL and RTL. Some of these involve bipolar transistors, others use field-effect devices. These other bipolar transistor gates, particularly, have the disadvantage that they are appreciably more complicated than DCTL or RTL. Hence each such gate occupies more real estate on the integrated-circuit silicon chip than, say, DCTL would. This greater area requirement is, of course, a disadvantage in LSI and even in MSI. As we shall see, FET devices are more economical of area but unfortunately are appreciably slower than bipolar transistor gates.

RTL and DCTL were the first logic families commercially developed. Thereafter a number of other logic families were introduced; these families are described and analyzed in succeeding chapters more or less in the order of their development. Integrated-injection logic (IIL or I²L) is the most recent logic system to be introduced to commercial application. We discuss it at this point because it has a special relationship to DCTL. IIL has the elegant simplicity of DCTL. A typical gate uses very little real estate and consumes very little power. For these reasons IIL is eminently suited for medium- and large-scale integration applications.

When a logic family is to be used in medium- or large-scale integration, where gates are to be crowded as close together as possible, the power dissipation per gate is a matter of great concern. It is generally true, as we have already seen for RTL, that a trade-off can be made by sacrificing power dissipation to speed. Hence a figure of merit which is relevant in comparing one logic family with another is the *speed-power product*. Speed is measured by the propagation time and power by the power dissipation of a typical gate. It is impressive to compare the speed-power product of IIL with the corresponding product for other logic families. By way of example, consider the comparison between IIL and TTL (transistor-transistor logic, discussed in detail in Chap. 6). At the present time TTL is the most popular family of logic, certainly in small-scale integration, and it has extensive applications in medium-scale integration and some application in large-scale integration. We find that for TTL the speed-power product [dimensionally, $(\text{time} \times \text{energy})/\text{time} = \text{energy}$] is typically 100 pJ, while for IIL this product is in the range 0.1 to 0.7 pJ. And while TTL gates can be packed with a density of about 20 gates per square millimeter, IIL gates allow a packing density in the range from 120 to 200 gates per square millimeter.

At the present writing (summer 1976) IIL logic is not commercially available in small-scale integration. Packages containing one or several gates, as available in RTL and in other logic families, have not been marketed. On the other hand, medium- and large-scale-integration chips are available.

Basic configuration of IIL The DCTL gate shown in Fig. 4.2-1 and the RTL gate shown in Fig. 4.1-1 do indeed look like gates. Each exhibits multiple inputs and a single output, as expected of a gate. (The other families of logic discussed in succeeding chapters also look like gates.) On the other

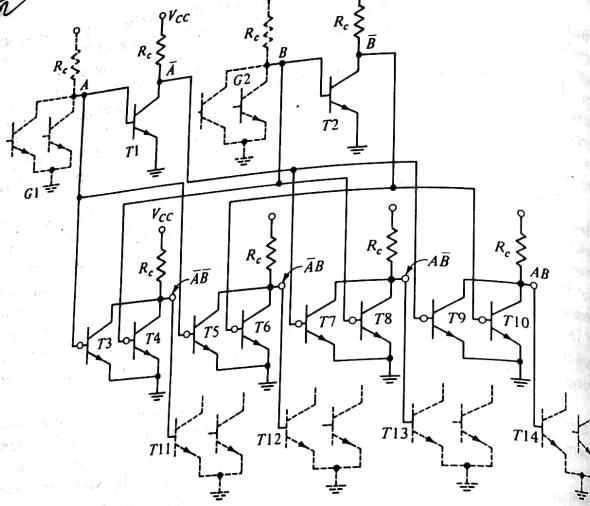


FIGURE 4.13-1
A DCTL gate structure which generates AB , $A\bar{B}$, $\bar{A}B$, and $\bar{A}\bar{B}$ from the logical variables A and B .

hand, the basic structure of an IIL exhibits a single input and multiple outputs, and some explanation is in order.

To pursue the matter, by way of illustration, let us consider that we have two logical variables to deal with, A and B , and that we need to generate the functions AB , $A\bar{B}$, $\bar{A}B$, and $\bar{A}\bar{B}$. We shall use DCTL logic. The variables A and B may themselves be functions of still other variables and hence will themselves be initially available as the outputs of other DCTL gates, $G1$ and $G2$, as shown in Fig. 4.13-1. Since these gates are external to the system we are to assemble, we have drawn them with dashed lines. (Two transistors are indicated in $G1$ and $G2$, but of course the number is arbitrary.)

We need \bar{A} and \bar{B} , and these functions of A and B are generated by the single-input gates involving transistors $T1$ and $T2$. Finally, with A , B , \bar{A} , and \bar{B} available, the functions required are generated by four two-input DCTL gates. The generated functions may themselves serve as inputs to other gates. To allow for this possibility we have connected the output $\bar{A}B$ to the base of $T11$, the output $\bar{A}\bar{B}$ to the base of $T12$, etc., these transistors, $T11$ through $T14$, providing one input of these other external gates.

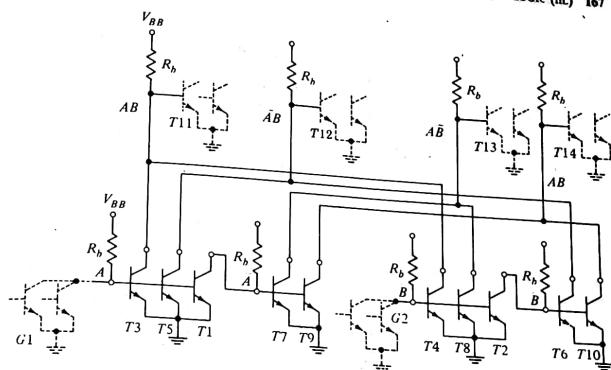


FIGURE 4.13-2
Figure 4.13-1 redrawn to group transistors with common base connections.

We have redrawn Fig. 4.13-1 to appear now as in Fig. 4.13-2. In redrawing that part of the figure of interest to us, i.e., the part drawn in solid lines, we have reorganized the drawing to group together transistors with common base connections. In Fig. 4.13-1 the transistors grouped together have common collector connections. Further, in going from Fig. 4.13-1 to Fig. 4.13-2 we have adopted a different interpretation of the function of the resistors. In Fig. 4.13-1 we view the resistors as common-collector resistors and accordingly label them R_c . In Fig. 4.13-2 the resistors are viewed as common-base resistors and are labeled R_b . Thus the resistor which was previously viewed as the common-collector resistor of the transistors of gate $G1$ is viewed instead as the common-base resistor of transistors $T1$, $T3$, and $T5$. Correspondingly, in one case the voltage to which the resistor has been returned is called V_{cc} and in the other it is called V_{bb} .

We have redrawn Fig. 4.13-2 to appear as in Fig. 4.13-3. Here groups of transistors with common emitters and bases (such as $T3$, $T5$, and $T1$) have been represented as a single multiple-collector transistor. The basic structure of the IIL gate is a transistor with multiple collectors and a single emitter together with a mechanism for supplying base current. The logic operation performed, as in DCTL, is the NOR operation. Thus, if a collector of a gate whose input is A is connected to a collector of a gate whose input is B and the joined collectors are in turn connected to the base of another gate, the logic that appears at the base is $\bar{A} + \bar{B}$.

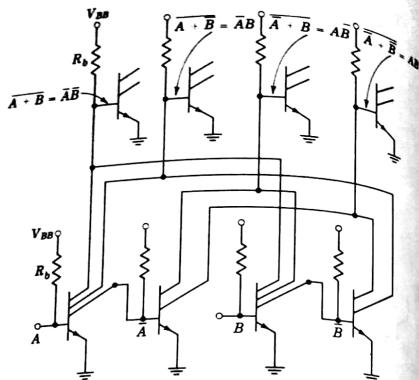


FIGURE 4.13-3
Figure 4.13-2 redrawn with multiple-collector transistors.

the base current of the multicollector transistors is supplied through a resistor. We now need to take note of the fact that actually this current is supplied through a transistor. Thus, finally, the basic configuration encountered in IIL appears as in Fig. 4.13-4. The transistor, it turns out, can be incorporated in the structure of the integrated circuit in a manner which uses much less chip area than would be required by a resistor. Observe that, for reasons to be discussed later, this added transistor is of the *pnp* type.

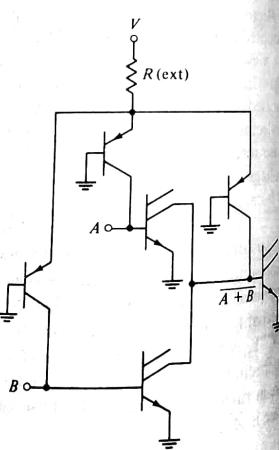


FIGURE 4.13-4
The basic configuration of IIL gates.

Up to this point we have made it appear that the base current of the multicollector transistors is supplied through a resistor. We now need to take note of the fact that actually this current is supplied through a transistor. Thus, finally, the basic configuration encountered in IIL appears as in Fig. 4.13-4. The transistor, it turns out, can be incorporated in the structure of the integrated circuit in a manner which uses much less chip area than would be required by a resistor. Observe that, for reasons to be discussed later, this added transistor is of the *pnp* type.

4.14 PHYSICAL LAYOUT OF IIL

In order to appreciate the features of IIL which account for its comparative merits, it is necessary to give some consideration to the physical construction of an IIL integrated circuit. We shall consider the matter in a simplified manner.

In integrated circuits, for reasons concerned with the technology of fabrication, transistors are of the *npn* type. The layout of an array of transistors is shown in Fig. 4.14-1. (In Fig. 4.14-1 for simplicity we have deliberately omitted a number of features needed for the proper operation of the device but not essential to our discussion.) The substrate is *p*-type silicon, and the *p*-type material extends upward from the substrate to the integrated-circuit surface to provide isolation from one transistor to the next. The collector, base, and emitter are *n*-type, *p*-type, and again *n*-type, respectively, as shown. Isolation is effected by arranging that the substrate shall be always negative with respect to the collector. The various parts of the transistors are formed by diffusions of impurities from the surface of the integrated circuit. The collector is relatively lightly doped, the base doped more heavily, and the emitter doped quite heavily.

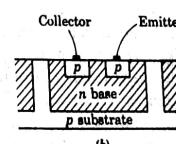
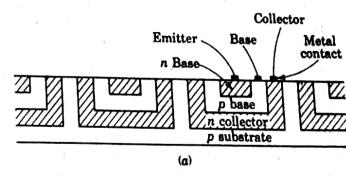


FIGURE 4.14-1
(a) The physical structure (simplified) of an integrated-circuit transistor. (b) The physical structure of a *pnp* lateral transistor.

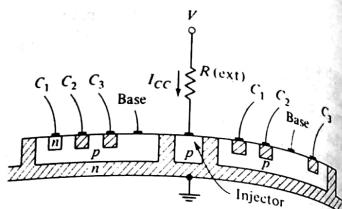


FIGURE 4.14-2
The physical structure of the basic gate configuration in IIL.

Both the geometry of the structure and the doping serve to ensure that the common-base gain α_N shall be close to unity and that $h_{FE} = \alpha_N/(1 - \alpha_N)$ shall be large. Since the emitter is doped much more heavily than the base, the carriers which cross the emitter junction are predominantly electrons. It is important that such be the case since holes crossing from base to emitter constitute an emitter current which will not be collected at the collector junction. Since the base is very thin, very few of the electrons injected into the base will be lost in the base as a result of recombination. Finally, we note that the emitter is almost surrounded by the collector, i.e., when viewed from the emitter, the collector subtends a large angle. As a consequence, an electron leaving the emitter, from no matter what point and in which direction, can hardly fail to reach the collector. For all these reasons we find $\alpha_N \approx 0.98$ or even higher. In the transistor configuration of Fig. 4.14-1a the axis of symmetry of current flow is in the vertical direction. Hence such a transistor is called a *vertical transistor*.

The transistor structure can be modified to allow a *pnp* transistor. Such a transistor is shown in Fig. 4.14-1b. The substrate is again *p* type. However the layer above the substrate becomes not the collector but the base. As shown, the emitter and collector are formed by *p*-type diffusions into the base. Observe that in the *pnp* case the angle subtended by the collector at the emitter is relatively small. For this reason the current gain h_{FE} of the *pnp* transistor is rather low, being in the range 0.5 to 5 compared with 50 to 150 for an *npn* transistor. Finally we note that the general direction of current flow is horizontal, i.e., to the side rather than vertical, hence the configuration of Fig. 4.14-1b, is called a *lateral transistor*.

Turning now to the *IIL* gate, we see the structure of such a gate in Fig. 4.14-2. (In this figure, we have again taken some liberties in the direction of simplification.) In this figure we display two three-collector gates side by side. The lower *n* layer, which in the conventional integrated-circuit transistor is the collector, is here the emitter. Since all the emitters in all the transistors operate at the same voltage (ground), a *p*-type isolation, necessary in Fig. 4.14-1a, is not required in the circuit of Fig. 4.14-2. In the conventional transistor these small *n* regions, embedded in the *p*-type base, is the emitter. In the *IIL* transistor these small *n* regions are the multiple collectors. However, in the *IIL* transistor

the relative doping is geometrically but not functionally the same as in the conventional transistor; i.e., while in the conventional transistor the emitter is most heavily doped and the collector least heavily doped, in the *IIL* transistor the situation is reversed. Hence the *IIL* transistor operates in a mode which we would describe as the *inverse mode* if we were dealing with a conventional transistor. Accordingly we expect that the current gain for the *IIL* transistor will be low. Such is indeed the case. Yet it turns out to be possible to get current gains as large as 5 and even higher. A current gain of this magnitude is entirely adequate; for, as can be seen in Fig. 4.13-1, each collector needs to provide for a fan-out of only 1.

As indicated in Fig. 4.14-2, the location of the base connection for the *npn* transistor is not the same from transistor to transistor. This feature is useful in that it allows base connections and collector regions to be located as required to accommodate the necessary interconnections.

Figure 4.14-2 also shows the *pnp* transistor which supplies base current to the *npn* transistor (see Fig. 4.13-4). The *p* region marked "injector" is the emitter of this transistor, its collector is the *p*-type base of the *npn* transistors, and its base is the *n*-type emitter of the *npn* transistor. Altogether the two transistors, one *npn* and one *pnp*, are formed with only four separate regions, the two transistors using two regions in common. For this reason *IIL* is also called *merged-transistor logic* (MTL).

Observe that the *pnp* transistor through which base current is injected into the *npn* transistor is pictured in Fig. 4.14-2 as serving two transistors on each side. Note also that the *pnp* transistors are *lateral* transistors. Current I_{cc} is supplied for the injector from a supply source V through an external resistor R . The voltage from the *p*-type injector region to ground is the voltage across a forward-biased junction and turns out in the present case to be about 0.85 V. Fortunately this injector-to-ground voltage tracks very well from injector to injector throughout the integrated-circuit chip. Hence it turns out to be possible to operate all the injectors in parallel so that all the injector current required by a chip can be supplied through a single external resistor.

Adjustability of speed We have seen that in RTL gates it is possible to sacrifice power dissipation for the sake of an improvement in speed. This trade-off is effected by changing the sizes of the resistors. Small resistors allow larger currents, so that capacitances can charge more rapidly albeit at the expense of greater power dissipation. What is to be noted here is that different speeds are associated with different integrated-circuit chips. In *IIL*, as we shall now see, it turns out that the trade-off between speed and power can be effected on a single chip by the simple expedient of changing the current I_{cc} injected into the chip.

Since the voltage at the injector is constant (≈ 0.85 V), the input power is proportional to the average input current I_{cc} . (On a large-scale-integration chip we may reasonably expect that a nominally fixed fraction of the transistors will be conducting at any one time and hence that I_{cc} will be rather constant.)

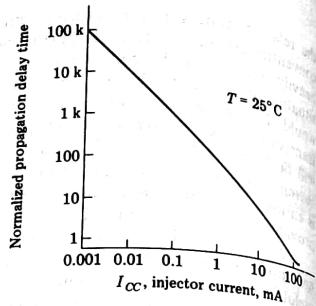


FIGURE 4.14-3
Normalized propagation delay time versus injector current.

A fraction I_{cc}/n of this current will be available, on the average, to operate a gate, n being the average number of gates which are conducting. At low current levels the propagation-delay time is principally the time required to charge junction and parasitic capacitors. The time required to charge these capacitors is inversely proportional to the available charging current I_{cc}/n . Accordingly the speed-power product is constant independently of I_{cc} . Increasing I_{cc} (by increasing V or decreasing R in Fig. 4.14-2), we can decrease the propagation delay at the expense of a proportionate increase in dissipation.

At medium current levels the principal source of propagation delay is the need to establish and remove the excess minority-carrier base charge in the transistors (see Sec. 1.17). This charge is proportional to the transistor current available to establish or remove it, and hence the propagation delay is independent of the current. Thus, in this current range, an increase in I_{cc} will increase the dissipation without reducing the delay.

At high current levels the transistors are driven into saturation. As discussed in Sec. 1.18, in saturation the stored base charge increases more than in proportion to transistor current. Hence, in this current range an increase in I_{cc} will not only increase the dissipation but will also increase the delay.

Figure 4.14-3 shows a plot of the normalized propagation delay time as a function of I_{cc} for the Texas Instrument (SBPO400) IIL chip. In the range of I_{cc} from 0.001 to 100 mA we can trade off dissipation and speed. Beyond about 100 mA an increase in dissipation yields no continuing decrease in speed. While the plot is not carried to the high-current region, at such high currents the plot would no doubt have a positive slope.

4.15 AN IIL DECODER

In Fig. 4.14-2 we show an injector serving two transistors. Actually a single injector can serve many transistors. When this is intended, the injector is

extended into a long strip parallel to the surface of the chip and is referred to as an *injector rail*. The transistors are fabricated on both sides of the rail and extend perpendicular to the rail.

As an example of an alternative layout we have the 3-bit decoder shown in Fig. 4.15-1. The function of decoders and their logic is discussed in Sec. 12.9. It will be sufficient here to note that the present decoder has as input the three logic variables A , B , and C and is intended to make available on eight separate output lines the eight minterms (see Sec. 3.17) of the input variables. Hence, at any time, only one output line will be at logic 1; all others will be at

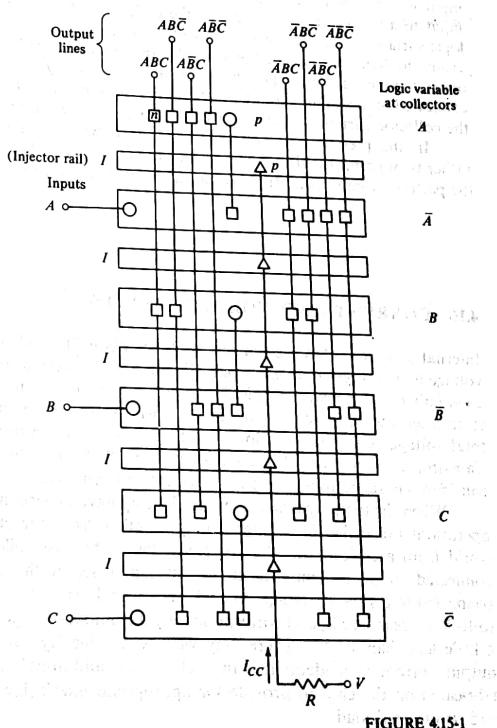


FIGURE 4.15-1
A 3-bit IIL decoder.

logic 0. The line selected to be at logic 1 is determined by the logic levels of the inputs.

The view shown is seen looking down at the surface of the chip. The common n-type emitter of the npn transistors (see Fig. 4.14-2) (which is also the base of the injector transistor) is not explicitly shown. We do see the p-type injector rail and the p-type base of the npn transistors (which is also the base of the injector transistor). Connections for the injected current are shown as triangles on the injector rail. The base connections are shown as circles and the collectors as squares. To the right of the figure are specified the logic variables present at the collector on each multicollector transistor. Thus the transistor whose input is A has \bar{A} at each collector. One of these collectors provides \bar{A} at the input to a second transistor, and at the collectors of this second transistor the logic variable is A . As we have already noted, the simple joining of collectors yields the logic product of the variables present at the collectors. It is to be understood, even though not shown in the figure, that each output line must be connected to the input of another gate so that the injector of that gate can provide the collector currents required for the transistors shown in the figure.

In the present case the transistors have been arranged to be parallel rather than perpendicular to the injector rails. Observe that, as a consequence, the pattern of interconnections is especially simple, requiring no crossovers.

4.16 CURRENT AND VOLTAGE LEVELS

Internal to an IIL chip the voltage levels are about 0.7 and 0.1 V. The higher voltage is measured at a base when all collectors connected to that base are not conducting. The lower level is the voltage at a collector and any base coupled to it when the collector is part of a transistor that is conducting. Thus the total voltage swing between logic levels is 0.6 V. The currents carried by the transistors are rather small in comparison with RTL and other gate types to be considered in succeeding chapters. Typically currents are in the range 1 to 10 μ A.

When the logic signals internal to the chip have undergone whatever logical operations are called for, a signal may be taken from the chip to the outside world from a collector of the last transistor. This last collector is not to be connected to a base and hence to develop a signal the collector must be connected to a supply voltage through an external resistor. Supply voltages up to 10 V are reasonable, and correspondingly the output-voltage swing will be only a little less than 10 V. There may well be a difficulty because the available output current is inadequate. In such a case additional transistors may be fabricated on the chip to provide for appropriate interfacing between the chip and its external load.

REFERENCES

- Hart, C. M., A. Slob, and H. E. J. Wulms: Bipolar LSI Takes a New Direction with Integrated Injection Logic, *Electronics*, Oct. 3, 1974, pp. 111-118.
- Berger, H. H., and S. K. Wiedmann: Merged-Transistor Logic (MTL): A Low-Cost Bipolar Logic Concept, *IEEE J. Solid-State Circuits*, October, 1972, pp. 340-346.
- Hart, K., and A. Slob: Integrated Injection Logic: A New Approach to LSI, *IEEE J. Solid-State Circuits*, October 1972, pp. 346-351.
- Alstein, J.: I²L: Today's Versatile Vehicle for Tomorrow's Custom LSI, *EDN*, Feb. 20, 1975, pp. 34-38.
- De Troge, N. C.: Integrated Injection Logic: Present and Future, *IEEE J. Solid-State Circuits*, October 1974, pp. 206-211.
- Horton, R. L., J. Englaide, and G. McGee: I²L Takes Bipolar Integration a Significant Step Forward, *Electronics*, Feb. 6, 1975, pp. 83-90.

5

DIODE-TRANSISTOR LOGIC

Diode-transistor logic (DTL) is a logic family in which diodes are used to provide the logic function. The circuit of Fig. 5.1-1 shows a DTL gate. Transistor T_0 is the output transistor (corresponding to T_2) of a preceding gate. Three inputs are indicated, but, of course, more can be added by adding more diodes to the array of diodes D_A , D_B , and D_C . The gate performs the NAND operation for positive logic. Consider that the driving gate T_0 is at logic level 0. Then T_0 is in saturation, and the corresponding collector voltage is $V_i \approx 0.2$ V or lower. We can verify that when the input in Fig. 5.1-1 is $V_i = 0.2$ V, T_2 is cut off and the output at the

In this chapter we consider a logic family, *diode-transistor logic* (DTL), whose circuitry is somewhat more involved than RTL. Although DTL has the advantage of greater fan-out and improved noise margins, it suffers from somewhat slower speed.

5.1 DIODE-TRANSISTOR-LOGIC (DTL) GATE

A diode-transistor-logic gate as realized with discrete components is shown in Fig. 5.1-1. Transistor T_0 is the output transistor (corresponding to T_2) of a preceding gate. Three inputs are indicated, but, of course, more can be added by adding more diodes to the array of diodes D_A , D_B , and D_C .

The gate performs the NAND operation for positive logic. Consider that the driving gate T_0 is at logic level 0. Then T_0 is in saturation, and the corresponding collector voltage is $V_i \approx 0.2$ V or lower. We can verify that when the input in Fig. 5.1-1 is $V_i = 0.2$ V, T_2 is cut off and the output at the

DIODE-TRANSISTOR LOGIC 177

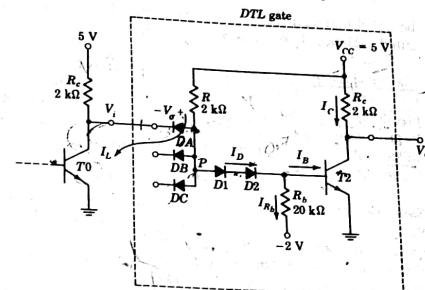


FIGURE 5.1-1
A discrete-circuit DTL gate.

collector is $V_{CC} = 5$ V, which corresponds to logic level 1. For, with the input at 0.2 V, the voltage at point P is

$$V_P = V_{CE(\text{sat}, T0)} + V_{DA} = 0.2 + 0.75 = 0.95 \quad \text{at } T = 25^\circ\text{C} \quad (5.1-1)$$

Here we have assumed that the current in diode D_A , which is approximately 2 mA, is adequate to produce a voltage drop of 0.75 V across diode D_A . The base-to-ground voltage of T_2 is then

$$V_{B2} = V_P - V_{D1} - V_{D2} = 0.95 - 0.65 - 0.65 = -0.35 \text{ V} \quad (5.1-2)$$

Here we have taken $V_{D1} = V_{D2} = 0.65$ since, as will appear, the currents in D_1 and D_2 are very small. This voltage V_{B2} is less than the cut-in voltage $V_i = 0.65$ V of transistor T_2 , and therefore T_2 is cut off. Thus, the current in diodes D_1 and D_2 continues through resistor R_b , and this current is

$$I_{R_b} = I_D = \frac{V_{B2} + 2}{R_b} = \frac{-0.35 + 2}{20 \times 10^3} = 0.08 \text{ mA} \quad (5.1-3)$$

This result confirms our initial assumption that the diode current is indeed quite small. Finally, since T_2 is cut off, the output voltage $V_o = 5$ V, corresponding to logic level 1.

We have shown that if V_i is at logic level 0, V_o is at logic level 1. Similarly we can show that if all of the inputs are at logic level 0, V_o remains at logic level 1. For, if all the inputs are simultaneously at logic level 0, the current through resistor R will divide among the diodes D_A , D_B , D_C , ... With a smaller current through, say, D_A , the drop across this diode will decrease slightly. It should

be noted that as a result V_p decreases slightly, so that T_2 is forced further into cutoff. Hence, V_o remains at the 1 level.

Current sinking The current I_L drawn out of diode D_A , as indicated in Fig. 5.1-1, flows into the collector of the driving saturated transistor T_0 . This current I_L has its source in the supply voltage V_{cc} . The process of returning this current to ground, which is also the negative return of the supply source, is commonly called *sinking*. Thus, transistor T_0 sinks the current drawn out of the input of the driven gate.

Saturation of T_2 If any one of the inputs is at logic level 0, T_2 is cut off and the gate output is at logic 1. If, however, all inputs are at logic level 1 (5 V), the current through R will flow through D_1 and D_2 and into the base of T_2 . Transistor T_2 will be driven into saturation, and the output will drop to logic level 0, as required for NAND operation.

To see that transistor T_2 does indeed become saturated let us calculate the base current $I_B = I_D - I_{R_b}$. In this case, $V_{B2} = 0.75$ V, and the current in R_b is

$$I_{R_b} \approx \frac{0.75 + 2}{20 \times 10^3} \approx 0.14 \text{ mA} \quad (5.1-4)$$

The voltage at point P is now $V_p = V_{D1} + V_{D2} + V_{B2} \approx 2.25$ V, where we have assumed that I_D and I_B are sufficiently large to cause the diode and base-emitter voltage to be approximately 0.75 V. Thus, the diode current I_D , which is also the current in resistor R , is

$$I_D = \frac{V_{cc} - V_p}{R} = \frac{5 - 2.25}{2 \times 10^3} \approx 1.4 \text{ mA} \quad (5.1-5)$$

The base current of transistor T_2 is the difference between the currents I_D and I_{R_b} , and is

$$I_B = I_D - I_{R_b} \approx 1.26 \text{ mA} \quad (5.1-6)$$

which is clearly sufficient to saturate transistor T_2 , since with $V_{ce}(\text{sat}) \approx 0.2$ V, $I_c = 2.4$ mA. Assuming that $h_{fe} = 50$, this collector and base current corresponds to $\sigma = 0.04$. Referring to Fig. 1.10-1, we find that $V_{ce}(\text{sat})$ is more nearly 0.1 V than 0.2 V.

Base resistor R_b If transistor T_2 is in saturation, and if then one or more of the gate inputs returns to logic level 0, point P falls to $V_p = 0.95$ V. The equivalent circuit of the gate at this instant is as shown in Fig. 5.1-2. Note that the voltage drop across both diodes D_1 and D_2 is 0.2 V and therefore these diodes are cut off. Hence, $I_D = 0$, and there is no current source to supply I_B .

If there were no charge storage in the system, I_B would instantly drop, causing the base-emitter voltage to fall below its cut-in voltage, and therefore T_2 would immediately cut off. The steady-state base-emitter voltage would then be -0.35 V [see Eq. (5.1-2)]. However, there is charge stored in the transistor.

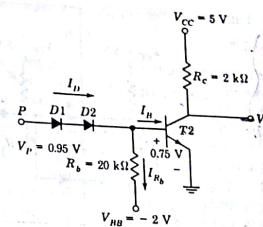


FIGURE 5.1-2
Equivalent circuit of the DTL gate at the instant that V_p is decreased to 0.95 V

When point P drops, momentarily cutting off diodes D_1 and D_2 , this stored charge leaves through resistor R_b . Thus, resistance R_b provides a discharge path for the charge stored in the transistor. Resistor R_b is connected to the -2 V supply to increase the rate of discharge.

The selection of $R_b = 20 \text{ k}\Omega$ and $V_{bb} = -2$ V (rather than, say, $R_b = 1 \text{ k}\Omega$ and $V_{bb} = -4$ V) is the result of a compromise. Many values of resistance and supply voltage are possible, and most will result in satisfactory operation. To increase the rate of charge removal and therefore decrease the time needed to cut off T_2 , we would like R_b to be small and V_{bb} to be very negative. However, when T_2 is cut off and we attempt to turn it on quickly, we would like all the diode current I_D to flow into the base of T_2 . In this case we would like R_b to be very large and V_{bb} to be positive. Typical values of R_b range from 5 to 30 kΩ, and values of V_{bb} vary from 0 to -5 V.

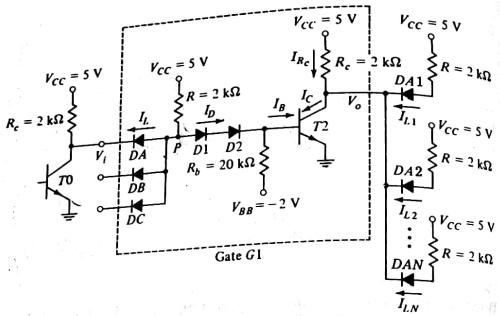
5.2 FAN-OUT

Figure 5.2-1 shows the DTL gate driving N other gates. When the gate output transistor T_0 is sinking the current of a gate G_1 that it is driving, it encounters its heaviest loading when all the other inputs of the driven gate are at logic level 1 (or equivalently when all other inputs are left floating). For, in this case all the current through R continues through D_A and into the collector of T_0 . On the basis of the previous discussion [see Eq. (5.1-1)] this current is

$$I_L = \frac{5 - 0.95}{R} = \frac{4.05}{2 \text{ k}\Omega} \approx 2 \text{ mA} \quad (5.2-1)$$

In any event, it is clear that to minimize the input loading due to DTL gate G_1 it is advantageous to make R a large resistance.

On the other hand, consider that all inputs of gate G_1 are at logic level 1. In this case current I_D flows from V_{cc} , through R and the diodes D_1 and D_2 , and finally divides between R_b and the base of T_2 . Transistor T_2 should now

FIGURE 5.2-1
A DTL gate with fan-out.

be in saturation. Now, to see the relationship between the fan-out N of gate G_1 and the resistor R , let us assume that the collector voltage of T_2 , when saturated, is $V_o = 0.2$ V. Then the current in each of the N loading gates is

$$I_{L1} = I_{L2} \cdots = I_{LN} \approx \frac{V_{CC} - V_{D41} - V_o}{R} = \frac{5 - 0.75 - 0.2}{R} \approx \frac{4}{R} \quad (5.2-2)$$

To arrive at this result we have assumed that all the current flowing in R , in each of the driven gates, flows into T_2 . We are therefore neglecting the current flow in diodes $D1$ and $D2$ of these driven gates. This neglected current is approximately 0.08 mA [see Eq. (5.1-3)] and is negligible.

The total collector current in T_2 is then

$$I_C = I_R + NI_{L1} = \frac{V_{CC} - V_o}{R} + \frac{4N}{R} = 2.4 \text{ mA} + \frac{4N}{R} \quad (5.2-3)$$

The base current in T_2 is I_B and is found, as before, by noting that $V_P = 2.25$ V when T_2 is saturated. Again, neglecting the current in R_b , we have

$$I_B \approx I_D = \frac{V_{CC} - V_P}{R} = \frac{2.75}{R} \quad (5.2-4)$$

Since we require that T_2 be driven to saturation, we also have

$$I_C = \sigma h_{FE} I_B \quad (5.2-5)$$

in which the value of σ can be estimated from Fig. 1.10-1 after we have decided what value we want for $V_{CE}(\text{sat})$, that is, how deeply T_2 is to be driven into saturation.

$$\begin{aligned} & 10^{6 \times \sigma} e^{IB} \\ & 6 \times \sigma e^{IB} = 2^{0.4 \times \frac{4N}{R}} \\ & N = \frac{5^2}{R} \end{aligned}$$

Combining Eqs (5.2-3) to (5.2-5) and solving for N , the fan-out, we find

$$N \approx 0.7\sigma h_{FE} - 0.6R \times 10^{-3} \quad (5.2-6)$$

Observe that to increase N we must increase σ ; that is, we must restrict the extent to which T_2 is driven into saturation. As we increase N , a point will be reached where $\sigma = 1$ and T_2 is no longer saturated. This situation develops because each additional load on the driving gate requires T_2 to sink an additional 2 mA without a compensating increase in base current in T_2 .

Assume $h_{FE} = 50$, $R = 2 \text{ k}\Omega$, and that $\sigma = 0.85$. This value of σ corresponds to a modest excursion into saturation and yields $V_{CE}(\text{sat}) = 0.2$ V (see Fig. 1.10-1). In this case the first term in Eq. (5.2-6) is $0.7\sigma h_{FE} = 30$, while the second term is $0.6(2) = 1.2$. Thus if we decide that the output transistor need not be moderately saturated, the fan-out $N = 29$. This result would not change appreciably if R were doubled or halved. Hence, we may neglect this second term and estimate that $N = 0.7\sigma h_{FE}$, depending only on h_{FE} and not on R . This relative independence of N from R is to be anticipated on the basis of the following consideration. If R changes, say decreases, the base current of T_2 increases. However, so that there is no net increase in allowable fan-out. It should be noted that if we decided that $V_{CE}(\text{sat}) = 0.1$ V under maximum fan-out, then $\sigma = 0.1$ and $N = 2$. Now the fan-out depends on $0.7h_{FE}$ and R .

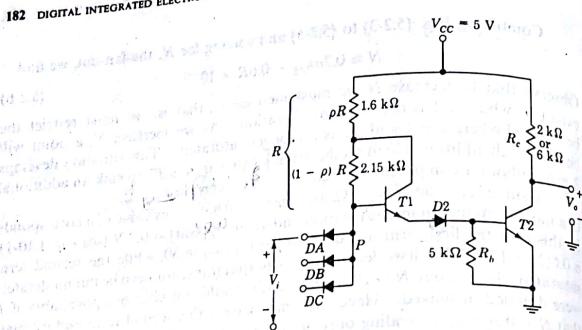
In the next section, when we consider the DTL gate in integrated-circuit form, we shall see how a modification of the circuitry devises to increase the base drive of the output transistor without a corresponding increase in the current load imposed by each additional gate. As a result the fan-out is significantly increased.

5.3 INTEGRATED-CIRCUIT DTL GATES

The DTL gate in integrated form is shown in Fig. 5.3-1. It is commercially available with either $R_c = 2 \text{ k}\Omega$ or $R_c = 6 \text{ k}\Omega$. In the former case the power dissipation is higher than in the latter case. However, all capacitance shunting the collector of T_2 to ground must charge through R_c when T_2 goes off. Hence, with $R_c = 2 \text{ k}\Omega$ the propagation delay time is smaller than with $R_c = 6 \text{ k}\Omega$.

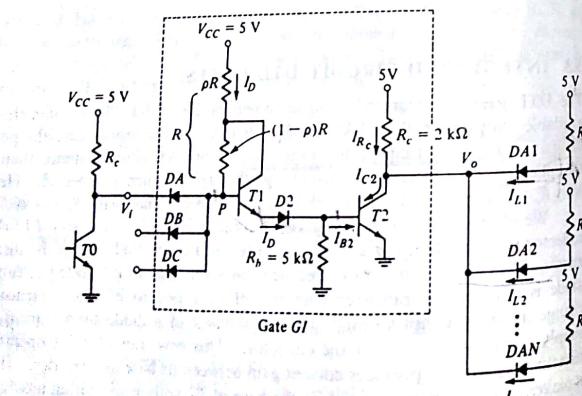
We observe that in the integrated circuit of Fig. 5.3-1 transistor T_1 , whose collector is connected to a tap on R , is used in lieu of diode $D1$ in the discrete circuit of Fig. 5.1-1. It is to be recalled, however, that in integrated circuitry a diode is normally a transistor with its collector tied to its base. Hence, the change involved is not so much the replacement of a diode by a transistor as simply a new connection for the collector. This new transistor T_1 operates in its active region and provides current gain between its base and emitter. Hence, it makes available more current for the base of T_2 without requiring a reduction in the resistance R .

We now explain the effect of this new transistor and the tap position. Let

FIGURE 5.3-1
An integrated-circuit DTL gate.

us begin by calculating the fan-out of the gate shown in Fig. 5.3-2. Let us assume, as before, that when T_2 is saturated, $V_o = 0.2$ V. Then, if the gate is fanned out to N loading gates, the load currents which must sink through T_2 are given, as in Eq. (5.2-2), as

$$I_{L1} = I_{L2} = \dots = I_{LN} \approx \frac{5 - 0.95}{R} \approx \frac{4}{R} \quad (5.3-1)$$

FIGURE 5.3-2
DTL gate with a fan-out of N .

Since the current in resistor R_c is $4/(2 \times 10^3) = 2.4$ mA, the collector current is

$$I_{C2} = 2.4 \times 10^{-3} + \frac{4N}{R} \quad (5.3-2)$$

This result, of course, is identical to that obtained in Eq. (5.2-3).

We now calculate the base current I_{B2} , again assuming that the current in R_b is negligibly small. First, we note that the voltage at point P is

$$V_p = V_{BE1} + V_{D2} + V_{BE2} = 0.75 + 0.75 + 0.75 = 2.25 \text{ V} \quad (5.3-3)$$

The calculation of $I_{B2} = I_D$ can now be easily performed with the help of Fig. 5.3-3. Let the base current in T_1 be called I . Then $I_{B2} = (h_{FE} + 1)I$.

$$V_{CC} - V_p = \rho R(h_{FE} + 1)I + (1 - \rho)RI \quad (5.3-4)$$

Solving for $(h_{FE} + 1)I = I_{B2}$ yields

$$I_{B2} = (h_{FE} + 1)I = \frac{V_{CC} - V_p}{R[\rho + (1 - \rho)/(h_{FE} + 1)]} \quad (5.3-5)$$

in which $V_{CC} - V_p = 2.75$ V.

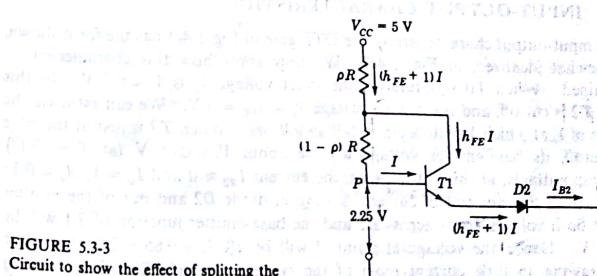
Since $\sigma h_{FE} I_{B2} = I_{C2}$, we have, combining (5.3-2) and (5.3-5),

$$\frac{2.75 \sigma h_{FE}}{R[\rho + (1 - \rho)/(h_{FE} + 1)]} = 2.4 \times 10^{-3} + \frac{4N}{R} \quad (5.3-6)$$

Solving for the fan-out N yields

$$N \approx \frac{0.7 \sigma h_{FE}}{\rho + (1 - \rho)/(h_{FE} + 1)} - 0.6R \times 10^{-3} \quad (5.3-7)$$

When $\rho = 1$, Eq. (5.3-7), as required, reduces to Eq. (5.2-6). As ρ decreases, the fan-out increases. In typical commercial DTL integrated-circuit gates $1/\rho$ is in the range 2 to 10. With $h_{FE} \approx 50$ and assuming, as in the previous section,

FIGURE 5.3-3
Circuit to show the effect of splitting the base resistor R .

that σ is in the range $\sigma \approx 0.85$, we have approximately that

$$N = \frac{0.7\alpha h_{FE}}{\rho} \quad (5.3-8)$$

Thus, we conclude that the new circuit increases the fan-out by the factor $1/\rho$ or, equivalently, that if we keep N fixed, we may drive the output transistor further into saturation by decreasing σ by the factor $1/\rho$.

The mechanism of increase in fan-out is the following. When the gate input V_i in Fig. 5.3-2 is at logic 0, T_1 is OFF and the driving gate sees the same load as in the circuit of Fig. 5.2-1. When, however, V_i is at logic 1, T_1 is ON and provides current gain. Hence the base current available for the output transistor is larger (by the factor $1/\rho$) than it would be if all the base current had to be furnished directly through the entire resistor R .

Equation (5.3-8) suggests that it might be advisable in Fig. 5.3-1 to move the tap on R closer to V_{CC} (letting $\rho \rightarrow 0$), thereby increasing the base drive of T_2 and the fan-out. The difficulty with such a procedure is that increasing the base current will result in an increase in the power dissipated in T_2 . Furthermore, if the gate is operated with a fan-out which is less than maximum, T_2 will be heavily saturated, thereby increasing the propagation delay time required to remove T_2 from saturation. The component values employed in the DTL gate represent a compromise, taking account of fan-out, propagation delay, and power dissipation.

A last feature to be noted in the gate on Fig. 5.3-1 is that the negative supply return of R_b has been eliminated. It is, of course, a great convenience to be able to operate the gate with a single supply voltage. We observe, however, that since we require that the stored charge in the base be drawn out of T_2 through R_b , the resistor R_b is much smaller in the integrated-circuit gate than in the discrete circuit.

5.4 INPUT-OUTPUT CHARACTERISTIC

The input-output characteristic of the DTL gate of Fig. 5.4-1 has the form shown, somewhat idealized, in Fig. 5.4-2. We now show how this characteristic is obtained. When T_0 is saturated, the input voltage V_i is $V_i \approx 0.2$ V. In this case T_2 is cut off, and the output voltage $V_o = V_{CC} = 5$ V. We can estimate the value of V_i at which V_o will begin to fall as follows. When T_2 is just at the edge of cutoff, its base-emitter voltage will be about $V_y = 0.65$ V (at $T = 25^\circ\text{C}$). Correspondingly, at this cut-in point, the current $I_{B2} \approx 0$ and $I_D = V_i/R_b \approx 0.13$ mA. With this amount of current flowing in diode D_2 and out of the emitter of T_1 both voltage drops across D_2 and the base-emitter junction of T_1 will be ≈ 0.7 V. Hence, the voltage at point P will be $2(0.7) + 0.65 = 2.05$ V. With through diode D_4 . It can be verified (Prob. 5.4-1) that at this point the current

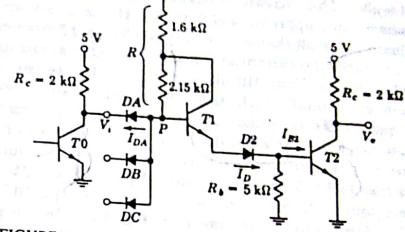


FIGURE 5.4-1
Circuit used to calculate the unloaded input-output characteristic.

$I_{DA} \approx 0.75$ mA. With this current, the drop across the input diode DA will be more nearly $V_{DA} \approx 0.75$ V. Hence (at $T = 25^\circ\text{C}$) the input voltage is $V_i = 2.05 - 0.75 = 1.3$ V, as indicated in Fig. 5.4-2.

We now calculate the minimum value of V_i needed to saturate T_2 . When T_2 switches from the edge of cutoff, $V_{AE2} = V_y = 0.65$ V, to saturation, $V_{CE2} = 0.2$ V, the collector current I_{C2} changes by approximately $\Delta I_{C2} \approx 2.4$ mA and the base current changes by $\Delta I_{B2} = \Delta I_{C2}/\sigma h_{FE} \approx 60$ μ A ($\sigma \approx 0.85$). Furthermore, since V_{AE2} changes by 0.1 V, from V_y to V_o , the change in the current

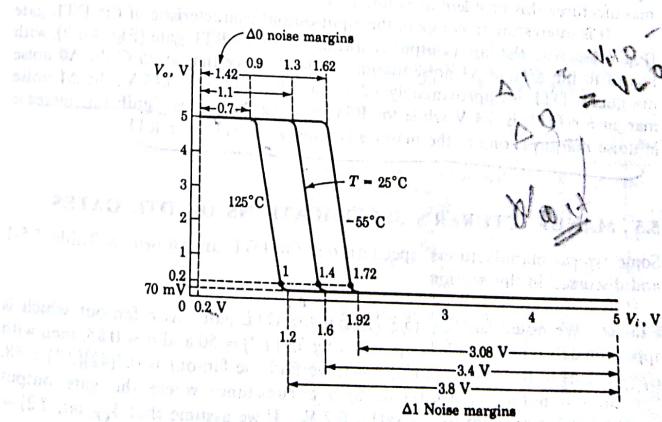


FIGURE 5.4-2
The unloaded input-output characteristic.

flowing in R_b is $\Delta I_{R_b} = 20 \mu\text{A}$. Hence, $\Delta I_{E_1} = 80 \mu\text{A}$ and $\Delta I_{S_1} \approx \Delta I_{E_1}/h_{FE} = 1.6 \mu\text{A}$. This increase in base current results in a decrease in diode current $I_{D_A} = 1.6 \mu\text{A}$, which is also approximately $1.6 \mu\text{A}$. Hence, a very small change in the current in diode D_A is all that is necessary to change $T2$ from cutoff to saturation.

Since the current in $T1$ and diode $D2$ has changed by $80 \mu\text{A}$, from $130 \mu\text{A}$ when $V_{BE2} = V_i$ to $210 \mu\text{A}$ when $V_{BE2} = V_o$, we can assume that V_{BE1} and V_{D_2} remain essentially unchanged. This assumption is certainly valid when considering diode D_A since the current in this diode changes by only $1.6 \mu\text{A}$. Thus, we can conclude that when the input voltage increases by 0.1 V from 1.3 to 1.4 V , the output transistor $T2$ moves from cutoff to saturation. When V_i increases further to 1.6 V , diode D_A cuts off (since $V_p = 2.25 \text{ V}$) and $I_{S_2} \approx 1.5 \text{ mA}$. At this point σ decreases to $\sigma \approx 0.03$ and $V_{CE(\text{sat})} \approx 70 \text{ mV}$ (see Fig. 1.10-1).

These results are shown in Fig. 5.4-2. In the plots for $T = -55^\circ\text{C}$ and $T = +125^\circ\text{C}$, also given in this figure, we have assumed, as usual, that the temperature sensitivity of a junction voltage is $-2 \text{ mV}/^\circ\text{C}$. However, as before, we have ignored the temperature dependence of $V_{CE(\text{sat})}$. The details of the calculations leading to the input-output characteristics are left as problems.

The input-output characteristic shown in Fig. 5.4-2 is somewhat idealized, inasmuch as we performed the calculation for the case where $T2$ is unloaded, i.e., has a fan-out of 0. A typical fan-out for a DTL gate is 8. If we had taken the fan-out into account, as illustrated in Fig. 5.3-2, we would see that when $T2$ is cut off, V_o is still equal to 5 V since $D41, D42, \dots, D4N$ are also cut off. As discussed in Sec. 5.3, the effect of too large a fan-out is to increase noise margin. For the maximum fan-out of 8 specified by the manufacturer this problem does not arise.

It is interesting to compare the input-output characteristic of the DTL gate (Fig. 5.4-2) with the input-output characteristic of the RTL gate (Fig. 4.6-2), with regard to the $\Delta 0$ and $\Delta 1$ noise margins. Here we see that at 25°C the $\Delta 0$ noise margin for DTL is approximately 1.2 V while for RTL it is 0.45 V ; the $\Delta 1$ noise margin for DTL is 3.4 V while for RTL it is 0.24 V . This significant increase in noise margin is one of the major advantages of DTL over RTL.

5.5 MANUFACTURER'S SPECIFICATIONS OF DTL GATES

Some typical manufacturers' specifications for DTL are shown in Table 5.5-1 and discussed in this section.

Fan-out We noted [see Eq. (5.3-7)] that the DTL gate has a fan-out which is approximately equal to $0.7h_{FE}/\rho$. Taking $h_{FE}(T2) = 50$ and $\sigma = 0.85$, then with $1/\rho = 2.3$, as in the circuit of Fig. 5.4-1, we find the fan-out is $0.7(42)(2.3) \approx 68$. This fan-out number corresponds to a circumstance where the gate output transistor $T2$ is brought to $V_{CE(\text{sat})} = 0.2 \text{ V}$. If we assume that $V_{CE(\text{sat}, T2)} = 0.1 \text{ V}$, then $\sigma = 0.1$ and the maximum fan-out is reduced to 8.

The manufacturer specifies a typical fan-out of 8 for the DTL gate and 25 for the buffer. The buffer employs an active pull-up similar to that employed in RTL. A detailed analysis of the DTL buffer is left for Prob. 5.4-6. Furthermore, each gate loading the driving gate appears as a $10-\text{pF}$ capacitor. Thus, a load of eight gates acts like a $80-\text{pF}$ capacitor and therefore affects the propagation delay time of the loaded gate. Thus, the specified fan-out of 8 ensures that the propagation delay time requirement of 30 to 40 ns is satisfied and that the collector-emitter voltage of the output transistor is approximately 0.1 V .

Voltage levels The specified voltage levels V_{oH} , V_{IH} , V_{IL} , and V_{oL} are shown in Fig. 5.5-1 for $T = 25^\circ\text{C}$. The voltage V_{oH} is the minimum output voltage of a gate corresponding to the logic 1 state. The value of the load current I_L when $V_o = V_{oH}$ is called I_{oH} . The value of 2.6 V shown in Table 5.5-1 is recommended by one manufacturer. Another manufacturer lists $V_{oH} = 4.3 \text{ V}$. To estimate V_{oH} we refer to Fig. 5.4-1 and let $T2$ be cut off. Then assuming $V_{CC} = 5 \text{ V} \pm 10 \text{ percent}$, $R_c = 2 \text{ k}\Omega$, a leakage current $I_L = 50 \mu\text{A}$, and $I_{oH} = -0.12 \text{ mA}$, we find

$$V_o = V_{oH} = 4.5 - 2000(50 + 120) \times 10^{-6} = 4.16 \text{ V} \quad (5.5-1)$$

If V_{oH} is calculated for $I_{oH} = 0$, that is, no load, we find

$$V_{oH} = 4.4 \text{ V} \quad (5.5-2)$$

This is similar to the value presented by one manufacturer.

Table 5.5-1 TYPICAL MANUFACTURERS SPECIFICATIONS AT 25°C FOR DTL GATE AND BUFFER

	Gate	Buffer
Output loading factor (fan-out)	8	25
Input loading factor	1	1
Power dissipation, mW	11	42
Input voltage, $V: V_{IL}$	1.1	1.1
V_{IL}	2	2
Output voltage, $V: V_{oL}$	0.4	0.4
V_{oL}	2.6*	2.6†
Output leakage current $I_L, \mu\text{A}$	50	50
Reverse diode current $I_R, \mu\text{A}$	2	2
Forward current of input diodes I_F for a unit input load, mA	1.6	1.6
Propagation delay time $t_{pd(HL)}$	30	40
$t_{pd(LH)}$	80	80

* $I_{oH} = -0.12 \text{ mA}$

† $I_{oH} = -2.5 \text{ mA}$.

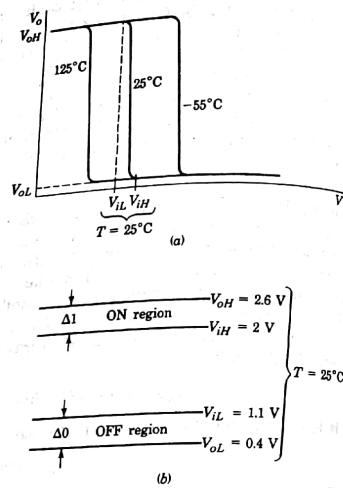


FIGURE 5.5-1
(a) Input-output characteristic defining worst case voltage levels. (b) Worst case voltage levels.

The maximum output voltage of the gate corresponding to the logic level 0 is V_{oL} . We have estimated that at a fan-out of 8, $V_{oL} = V_{CE(\text{sat})} = 0.1\text{ V}$. The manufacturer, in order to be extremely conservative, will only say that the collector-emitter voltage V_{oL} will never exceed 0.4 V.

The maximum allowable input voltage which will reliably be recognized as corresponding to logic level 0 is V_{iL} , and the minimum allowable input voltage which will reliably be recognized as corresponding to logic level 1 is V_{iH} . The temperature dependence of these parameters is brought out in Fig. 5.5-1. Our calculations indicate that, at 25°C , $V_{iL}(\text{max})$ should be approximately 1.3 V and $V_{iH}(\text{min})$ about 1.6 V. We find (see Table 5.5-2) that the manufacturer's

Table 5.5-2

	-55°C	25°C	125°C
V_{iH}, V	2.1	2	1.9
V_{iL}, V	1.4	1.1	0.8
$\Delta 0$	1.0	0.7	0.4
$\Delta 1$	0.5	0.6	0.7

specification on V_{iL} is 1.1 V, while the specification on V_{iH} is 2 V. Again, we note that these specified values represent worst-case conditions.

Noise immunity The $\Delta 0$ and the $\Delta 1$ noise immunity are given by

$$\Delta 0 = V_{iL} - V_{oL} \quad (5.5-3)$$

$$\Delta 1 = V_{oH} - V_{iH} \quad (5.5-4)$$

The parameters V_{iH} and V_{iL} as specified by the manufacturer are summarized in Table 5.5-2. Using $V_{oH} = 2.6\text{ V}$ and $V_{oL} = 0.4\text{ V}$ (since these values do not change significantly with temperature), we can calculate $\Delta 0$ and $\Delta 1$, which are also tabulated in Table 5.5-2. The value $V_{oH} = 2.6\text{ V}$ seems unreasonably conservative. It may have been selected by the one manufacturer in order simply to arrange that the noise margins $\Delta 0$ and $\Delta 1$ be nearly equal to one another. Note the significant increase in the noise immunity of DTL compared with that of RTL.

Propagation delays Propagation delays in DTL gates are of the order of 30 to 80 ns. The delay associated with turning on the output transistor [the turn-on delay $t_{pd}(HL)$] is smaller than the delay associated with driving that transistor back to cutoff [the turn-off delay $t_{pd}(LH)$]. The turnoff delay is generally substantially larger than the turn-on delay, often by a factor of 2 or 3. For, at turn-on, any capacitance shunting to ground the output of the gate can discharge rapidly through the low impedance of a transistor in saturation. At turnoff, however, this shunt capacitor must charge through the relatively large pull-up resistor R_c . In addition, at turnoff, there is a storage-time delay in the output transistor T_2 which is not encountered at turn-on.

5.6 THE WIRED-AND CONNECTION

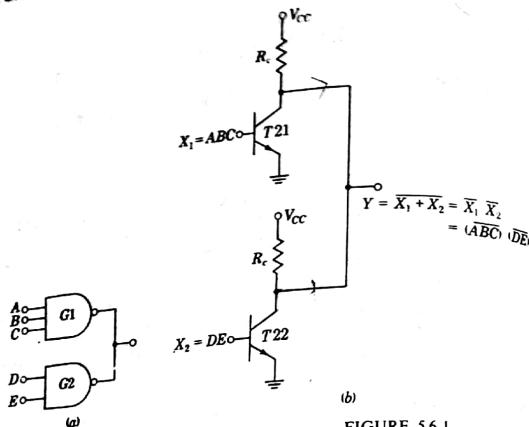
A useful extension of the logic capability of DTL gates can be achieved by joining the outputs of two DTL gates. Such a connection of outputs into a single common output is indicated in Fig. 5.6-1a. In Fig. 5.6-1b is shown the output transistor of each of the two gates with their collectors tied together to a common output Y . For the inputs given in Fig. 5.6-1a the logic variables which appear at the bases of these two transistors are $X_1 = ABC$ and $X_2 = DE$, as shown in Fig. 5.6-1b. Additionally, it is readily verified (compare with the RTL gate of Fig. 4.4-1) that $Y = \overline{X_1 + X_2}$. Hence, altogether we have

$$Y = \overline{X_1 + X_2} = \overline{ABC + DE} \quad (5.6-1)$$

or, using De Morgan's theorem,

$$Y = \overline{X_1} \overline{X_2} = (\overline{ABC})(\overline{DE}) \quad (5.6-2)$$

From Eq. (5.6-1) it appears as though tying the outputs together resulted in a NOR operation. On the other hand, from Eq. (5.6-2) we can interpret the result

FIGURE 5.6-1
The WIRED-AND connection.

as what would be obtained if the outputs of the individual DTL gates, when not joined, were combined in an AND gate. This result applies as well if more than two gates are connected at their outputs. This consideration leads to the description of the arrangement of Fig. 5.6-1 as a WIRED-OR or a WIRED-AND connection of DTL gates. In this text we shall use the terminology WIRED-AND.

To appreciate the usefulness of the WIRED-AND connection, consider that we did indeed want to generate the logical function Y given in Eq. (5.6-1) or (5.6-2); and suppose that we were restricted to using the DTL NAND gates in the conventional manner. Then, as already noted, we should have to combine the DTL-gate outputs \overline{ABC} and \overline{DE} in an AND gate. An AND gate, when constructed from NAND gates, requires two such NAND gates. Hence, altogether, the equivalent of the gating circuit of Fig. 5.6-1a would appear as shown in Fig. 5.6-2 and require a total of four gates.

Loading rules In the WIRED-AND connection of a number of DTL gates, it may happen that only one output transistor is conducting while the others are

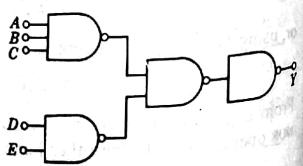


FIGURE 5.6-2

A NAND gate system to obtain $Y = (ABC̄)(DĒ)$.

all cut off. Then this transistor must not only sink the current of the loading gates and the current due to its own pull-up resistor but must also sink the current in the pull-up resistor of the other output transistors. To allow for this situation it is necessary to reduce the allowable fan-out of each gate in the WIRED-AND connection. This fan-out reduction is calculated in the following example.

EXAMPLE 5.6-1 A number K of DTL gates are connected in a WIRED-AND configuration. Calculate the reduction required in the output loading factor as a function of K .

SOLUTION Refer to Fig. 5.6-3. Here we see the outputs T_1, T_2, \dots, T_K of K gates connected in a WIRED-AND configuration and driving N DTL gates. The driven gates are represented by their input diodes and series resistors R since the load affects the operation of the driving gate only when the driving gate is saturated, in which case all the current in R flows through its series diode.

Now assume that X_1 is in logic level 1 while X_2, \dots, X_K are in logic level 0. Then T_1 is saturated. Let $V_{CE}(\text{sat}) \approx 0.2$ V. If the current in each diode, D_1, D_2, \dots, D_N , is called I_L , we have

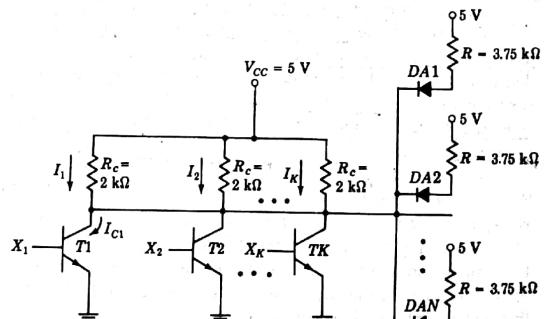
$$I_L = \frac{V_{CC} - V_D - V_{CE}(\text{sat})}{R} = \frac{5 - 0.75 - 0.2}{3.75 \times 10^3} = 1.08 \text{ mA} \quad (5.6-3)$$

The currents flowing in the collector resistors of the WIRED-AND output transistors are each equal to I_L , which is

$$I_1 = I_2 = \dots = I_K = \frac{V_{CC} - V_{CE}(\text{sat})}{R_c} = \frac{5 - 0.2}{2 \times 10^3} = 2.4 \text{ mA} \quad (5.6-4)$$

Thus, the collector current in T_1 is

$$I_{C1} = KI_1 + NI_L = 2.4K + 1.08N \text{ mA} \quad (5.6-5)$$

FIGURE 5.6-3
The WIRED-AND connection for K gates.

We see from this expression that increasing K by 1 gate is equivalent, as far as the increase in collector current is concerned, to increasing the fan-out by $\Delta N = 2.4/1.08 = 2.2$ gates. Thus, for each gate connected in parallel with T_1 to form a WIRED AND, we must reduce the maximum output loading factor by 2.2 gates. This results in leaving the collector current fixed, and therefore T_1 remains saturated. The manufacturer specifies a load reduction of 2.5 gates.

It is of great practical importance to note that DTL gates using an active pull-up, as in the buffer shown in Fig. P5.4-6, should not be used in the WIRED-AND mode. The reason for this restriction is made clear in Sec. 6.14.

5.7 HIGH-THRESHOLD LOGIC (HTL)

There are circumstances where logic circuits must operate in environments which are very noisy electrically. For operation in such surroundings there is available a line of DTL logic circuits with thresholds, i.e., noise immunities $\Delta 0$ and $\Delta 1$, which are quite high in comparison with the thresholds of conventional DTL circuitry.

A high-threshold-logic (HTL) gate is shown in Fig. 5.7-1. Comparing Fig. 5.7-1 with the conventional gate of Fig. 5.4-1, we note that in the HTL gate the supply voltage has been raised from 5 to 15 V. This feature, as we shall see, accounts for the increased noise immunity. Because of the higher supply voltage, the diode D_2 in Fig. 5.3-1 must sustain a higher voltage and has hence been replaced by a zener diode. We note, additionally, that in the HTL gate the resistance values are appreciably larger than in the conventional gate. This increase in the resistance values is necessary because of the increased supply voltage; for, otherwise, the increased supply voltage would result in a large increase in current and therefore in power dissipation in the HTL gate.

The use of these larger resistance values has an adverse affect on the speed of operation of the HTL gate. For now, when capacitances need to charge or discharge through these resistances, they do so in circuits having relatively larger time constants. Thus, while conventional DTL gates have propagation delay times which are, typically, some tens of nanoseconds, HTL gates have propagation times which may be as high as hundreds of nanoseconds.

The operating principles and characteristics of zener diodes are discussed in Sec. 1.4. The zener diode in the HTL gate of Fig. 5.7-1 has an operating voltage of about 6.9 V. Further, at this voltage the temperature sensitivity of the operating voltage of a zener diode is about comparable in magnitude to the temperature sensitivity of a forward-biased diode. However, the temperature sensitivity of the zener diode is positive, while the temperature sensitivity of a forward-biased junction is negative.

In the conventional DTL gate we noted, as in Fig. 5.4-2, that the input voltage at which the gate makes its transition between logic levels is dependent

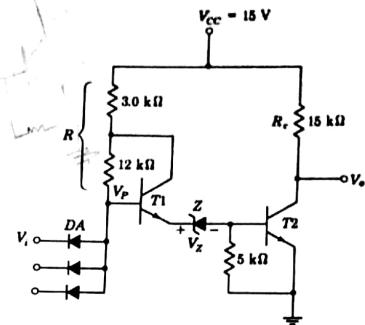


FIGURE 5.7-1
An HTL NAND gate.

on temperature. The temperature sensitivity of this voltage is equal to the temperature sensitivity of two diodes in series. For, in Fig. 5.4-1, since the input diode and the base-emitter junction of T_1 are in polarity opposition, the temperature sensitivities of these two junctions cancel. We are then left with the combined temperature sensitivities of D_2 and the base-emitter junction of T_2 . In the HTL gate of Fig. 5.7-1 the input diode and the junction of T_1 cancel, as before. But now we also have a cancellation of the temperature sensitivities of the zener diode and the junction in T_2 since the temperature sensitivities of these two are approximately equal and are also in opposite directions. The result is that the temperature sensitivity of the HTL gate is significantly less than that indicated in Fig. 5.4-2 for the DTL gate. We therefore ignore temperature effects.

5.8 INPUT-OUTPUT CHARACTERISTIC OF THE HTL GATE

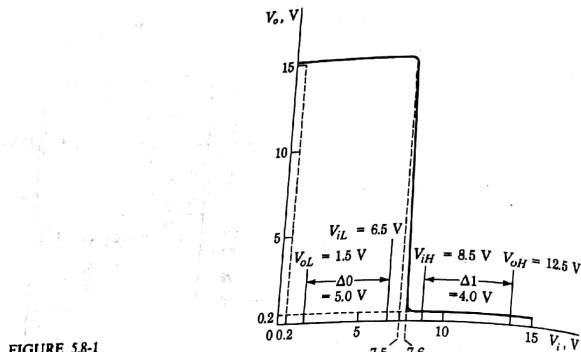
The input voltage at which transistor T_2 begins to come out of cutoff is (see Prob. 5.8-1)

$$\begin{aligned} V_i &= V_{BE2} + V_Z + V_{BE1} - V_{DA} \\ &= 0.65 + 6.9 + 0.7 - 0.75 = 7.5 \text{ V} \end{aligned} \quad (5.8-1)$$

The input voltage at which transistor T_2 is at the edge of saturation is (see Prob. 5.8-2)

$$\begin{aligned} V_i &= V_{BE2} + V_Z + V_{BE1} - V_{DA} \\ &= 0.75 + 6.9 + 0.7 - 0.75 = 7.6 \end{aligned} \quad (5.8-2)$$

Using the results in Eqs. (5.8-1) and (5.8-2), we can draw the input-output characteristic shown in Fig. 5.8-1.

FIGURE 5.8-1
Input-output characteristic of HTL.

5.9 MANUFACTURER'S SPECIFICATIONS

Manufacturer's voltage specifications for HTL are typically

$$\begin{aligned} V_{IL} &= 6.5 \text{ V} & V_{oL} &= 1.5 \text{ V} & I_{oL} &= 12 \text{ mA} \\ V_{IH} &= 8.5 \text{ V} & V_{oH} &= 12.5 \text{ V} & I_{oH} &= -30 \mu\text{A} \end{aligned} \quad (5.9-1)$$

These voltages have been marked on the plot of Fig. 5.8-1. The voltage range V_{IL} to V_{IH} bridges the transition region rather symmetrically and leaves a reasonable margin of safety. The voltage $V_{oH} = 12.5 \text{ V}$ is also rather reasonable. For, with a 15-kΩ pull-up resistor, as in Fig. 5.7-1, a leakage current in the output transistor of 100 μA, and with $I_{oH} = -30 \mu\text{A}$, the output voltage of the gate is $V_o = V_{CC} - 15 \times 10^3 \times (130 \times 10^{-6})$. Letting $V_{CC} = 14 \text{ V}$ (the manufacturer assumes that V_{CC} can vary by $\pm 1 \text{ V}$ from the nominal value of 15 V), we have $V_o = 12 \text{ V}$, which is actually somewhat less than the value of V_{oH} given.

The voltage $V_{oL} = 1.5 \text{ V}$ seems artificially high. The manufacturer states that even if 12 mA enters the output transistor T_2 from an outside source, $V_o \leq 1.5 \text{ V}$. Presumably, if a gate is operating properly, the output transistor should be in saturation and the gate output should be 0.2 or at most, 0.4 V (see Fig. 1.10-1). This value of V_{oL} is specified, we may surmise, to make the noise immunities $\Delta 0$ and $\Delta 1$ more nearly comparable. For we find that the noise margins $\Delta 1$ and $\Delta 0$ are

$$\Delta 1 = V_{oH} - V_{IH} = 12.5 - 8.5 = 4.0 \text{ V} \quad (5.9-2)$$

$$\Delta 0 = V_{IL} - V_{oL} = 6.5 - 1.5 = 5.0 \text{ V} \quad (5.9-3)$$

and

Propagation delay time The propagation delay time of HTL gates is typically

$$t_{pd}(LH) = 200 \text{ ns} \quad t_{pd}(HL) = 100 \text{ ns}$$

which is approximately 3 times as large as the delay found when using DTL.

Fan-out Manufacturers typically specify a fan-out of 10 for the HTL gate. As we shall now see, this figure is, as usual, conservative.

The derivation of fan-out for the DTL gate which leads to Eq. (5.3-7) applies to the present case as well. Using, in the derivation $V_{CC} = 15 \text{ V}$, $R = 15 \text{k}\Omega$ and $V_z = 6.9 \text{ V}$, we find that Eq. (5.3-7) becomes

$$N = \frac{0.5\sigma h_{FE}}{\rho + (1 - \rho)/(h_{FE} + 1)} - 0.07R \times 10^{-3} \quad (5.9-4)$$

In the HTL gate, $\rho = 0.2$, so that $N \approx 2.5\sigma h_{FE}$. If we use $h_{FE} = 50$ and $\sigma = 0.85$, which brings the output transistor just slightly into saturation, where $V_{CE(\text{sat})} = 0.2 \text{ V}$, we find $N = 106$. If we require that the output transistor be driven well into saturation [$\sigma = 0.1$, $V_{CE(\text{sat})} = 0.1 \text{ V}$], we find $N \approx 12$. Thus, $N = 10$ seems conservative yet reasonable.

REFERENCES

- Bohn, R., and R. Seeds: Collector Tap Improves Logic Gating, *Electronic Design*, August 3, 1964, pp. 51-55.

6

TRANSISTOR-TRANSISTOR LOGIC

TRANSISTOR-TRANSISTOR LOGIC 197

Transistor-transistor logic (TTL) is a form of digital circuitry which uses transistors as the basic switching elements. It is a relatively slow form of logic, but it has the advantage of being able to operate directly from a standard 5-volt power source. The TTL gate is a three-transistor circuit consisting of an input stage, a driver stage, and an output stage. The input stage consists of a single NPN transistor, T1, connected as an emitter-follower. The driver stage consists of two NPN transistors, T2 and T3, connected in a common-emitter configuration. Transistor T2 is controlled by the output of T1, and transistor T3 is controlled by the output of T2. The output stage consists of a single NPN transistor, T3, connected as an emitter-follower. The output voltage, V_o , is taken from the collector of T3. The TTL gate is a single-input gate, and its output is a single-bit digital signal.

6.1 TRANSISTOR-TRANSISTOR LOGIC (TTL)

The usefulness of a DTL gate is limited by its speed of operation. A principal source of this limitation can be appreciated by considering the circuit of Fig. 6.1-1, where a single-input gate is shown. With a view toward the TTL gate shortly to be introduced, we have explicitly shown the input diode as a diode-connected transistor, i.e., a transistor with collector and base connected.

Consider, now, that the input to the gate is at logic level 1. Then current flows through T_2 and diode D and into the base of T_3 . The transistor T_3 is driven to saturation, and the gate output is at logic level 0. Now change the input to logic level 0. Then the output of the gate should go to logic level 1. However, this transition to logic level 1 will not take place until transistor T_3 comes out of saturation, passes through the active region, and eventually goes to cutoff. Cutoff however, will not be reached until the stored base charge of T_3 has been removed (see Sec. 1.20). During this removal of the stored base charge, transistor T_1 is on, but transistor T_2 and diode D are cut off. Hence, there is no alternative: the base charge must leak off through the resistor R_b ,

or dissipate by recombination. This relatively slow mechanism for the removal of stored charge establishes the essential speed limitation of the DTL gate.

The DTL speed limitation is overcome in the transistor-transistor-logic gate (TTL). In its simplest and most elemental form this gate appears as shown in Fig. 6.1-2. (For the moment, we picture a single input gate.) There is a certain similarity between the TTL gate of Fig. 6.1-2 and the DTL gate of Fig. 6.1-1. In the TTL gate the input transistor T_1 is connected as a transistor, the collector-base connection being removed. Transistor T_2 and diode D of the DTL gate are removed, and the collector of T_1 is connected directly to the base of T_3 .

When, in the TTL gate, the input is high, the emitter-base junction of T_1 will be back-biased and current will flow, through R and through the forward-biased base-collector junction of T_1 into the base of T_3 . In this mode of

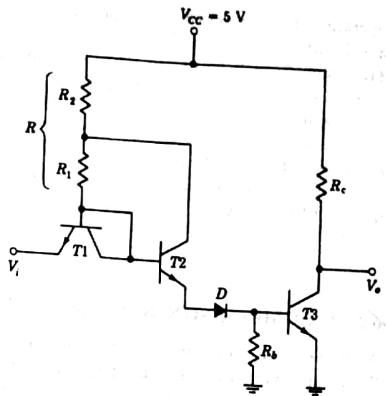


FIGURE 6.1-1
The basic DTL gate.

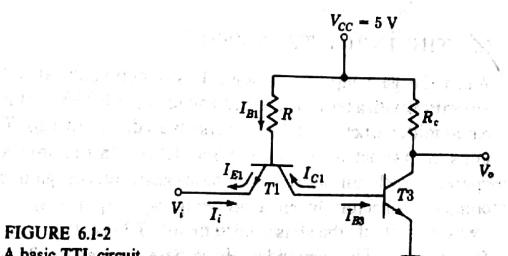


FIGURE 6.1-2
A basic TTL circuit.

operation the collector of transistor T_1 operates as an *emitter* and the *emitter* as a *collector*. Transistor T_1 is operating in the *inverse mode* (see Sec. 1.10). Transistor T_3 will be driven into saturation, and the gate output will be low (logic level 0). Now let the input drop to logic level 1. The emitter junction of T_1 will become forward-biased. We note that the base of T_3 is connected to the collector terminal of T_1 . The stored charge in the base of T_3 no longer leaks off through a resistor, as in the DTL gate, but flows out through the collector of transistor T_1 . Note that in the TTL gate, a base resistor for T_3 has not even been included. A simple comparative calculation is made in the next section.

6.2 A COMPARISON BETWEEN TTL AND DTL

A simple comparison between the DTL gate and the TTL gate will illustrate the great advantage of the latter in the speed of removal of stored base charge from T_3 . When T_3 is in saturation, its base-emitter voltage is $V_\sigma = 0.75$ V. When, in Fig. 6.1-1, diode D is cut off, the initial current out of the base of T_3 is $V_\sigma/R_b = 0.75/2 \text{ k}\Omega \approx 0.38 \text{ mA}$.

On the other hand, in the TTL gate, consider that the input is grounded. Then the base-emitter junction of T_1 is forward-biased, and we shall assume that the voltage across the junction is 0.75 V. As we shall see, in a typical case the resistor R in Fig. 6.1-2 is $R = 4 \text{ k}\Omega$. Then the base current in T_1 is $I_{B1} = (5 - 0.75)/4 \text{ k}\Omega \approx 1.1 \text{ mA}$. Initially, transistor T_1 is operating in its active region, since the base-to-ground voltage of T_3 , which is also the collector-to-ground voltage of T_1 , is initially $V_\sigma = 0.75$ V and therefore $V_{CE1} = 0.75$ V. We then observe that the initial collector current in T_1 , which is also the rate at which the stored base charge of T_3 is being removed, is $h_{FE} I_{B1}$, where h_{FE} is the current gain of T_1 . Even if we allow an h_{FE} no larger than $h_{FE} = 20$, we find the discharge rate to be $h_{FE} I_{B1} = 20(1.1) = 22 \text{ mA}$, which is to be compared with 0.38 mA in the DTL case.

This faster removal of the charge stored in T_3 results in TTL gates which operate at propagation delay times that are one-tenth those of DTL gates.

6.3 THE INPUT TRANSISTOR

When the gate input is at logic 1 (say nominally at 5 V), transistor T_3 is in saturation with a base-to-ground voltage $V_\sigma = 0.75$ V. It is then apparent that the base-emitter junction of T_1 is reverse-biased. Transistor T_1 is therefore operating in its inverse active region (see Sec. 1.10). In this inverse region the transistor operates with an inverse common-base current gain α_i , the corresponding common-collector current gain being $h_{FC} = \alpha_i/(1 - \alpha_i)$. Referring to Fig. 6.1-2, we see that if the base current of T_3 is I_{B3} , the input current I_{I1} is $I_{I1} = \alpha_i I_{B3}$. The remainder of the base current for T_3 is supplied by the base

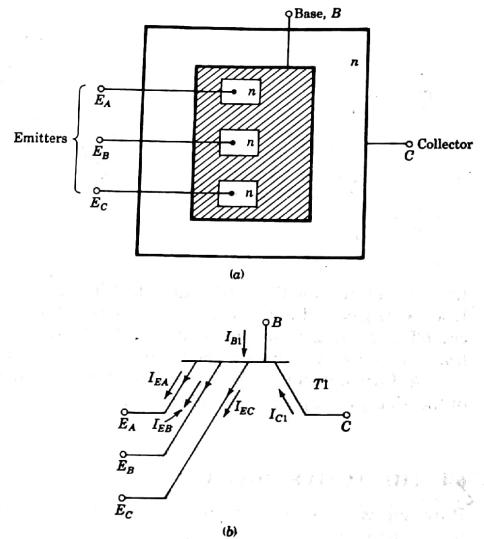
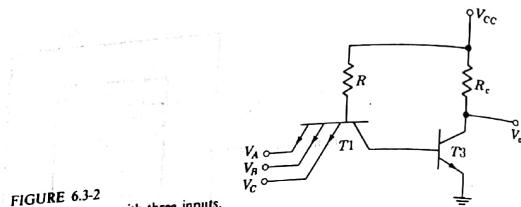


FIGURE 6.3-1
(a) Pictorial representation and (b) symbol of multiemitter transistor.

current I_{B1} of T_1 , so that $I_{B1} = (1 - \alpha_i)I_{B3}$. In a TTL gate, the input transistor T_1 is deliberately designed to have a very low value of inverse current gain. Values in the range $\alpha_i \approx 0.02$ or even lower are typical of TTL gates. With such a value of α_i , only 2 percent of the required base current of T_3 need be supplied by the input driving source, while the other 98 percent is supplied through R from V_{CC} . Thus, the low value of α_i has the advantage of minimizing the loading on a driving source, at least when the input is at logic level 1.

For the circuit of Fig. 6.1-2 to serve as a gate, additional inputs must be provided. These additional inputs may be made available by paralleling T_1 with additional input transistors. All such input transistors would then have their collectors tied together and their bases similarly joined. In practice we do not parallel input transistors but construct a transistor with a single common collector, a single common base, and multiple emitters. The physical structure of such a *multiple-emitter transistor* is shown in Fig. 6.3-1a, and its circuit symbol is shown in Fig. 6.3-1b. A three-input TTL gate using such a multiple-emitter

FIGURE 6.3-2
A basic TTL gate with three inputs.

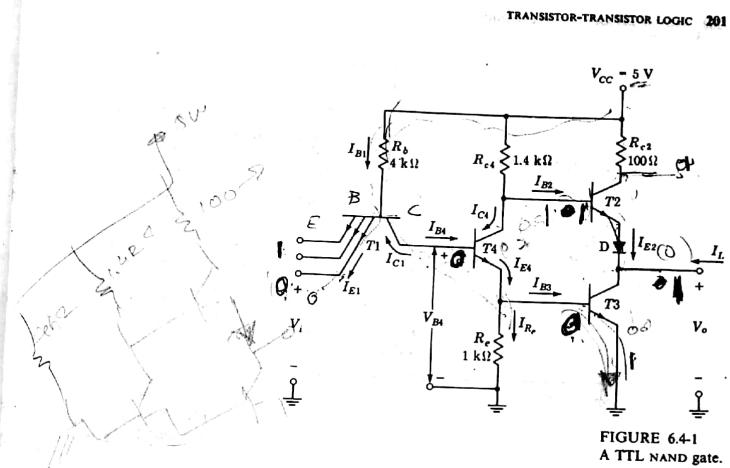
transistor is shown in Fig. 6.3-2. Like the DTL gate, the TTL gate of Fig. 6.3-2 is a NAND gate. If any one of the inputs is at logic level 0, transistor T_3 is cut off and the gate output is at level 1. If all the inputs are at logic 1, transistor T_3 is in saturation and the output is at logic 0.

A discussion of the effect of the multiple-emitter transistor on the operation of the TTL gate is postponed until Sec. 6.7.

6.4 THE ACTIVE PULL-UP

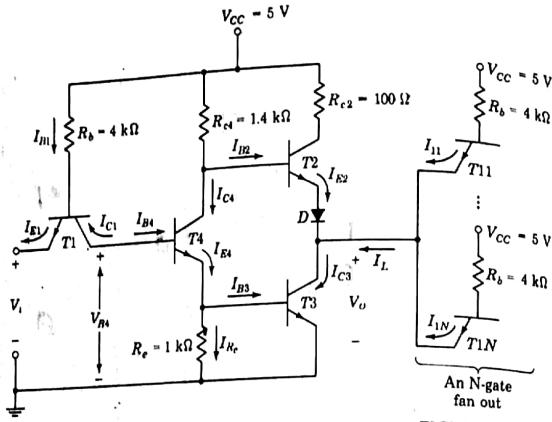
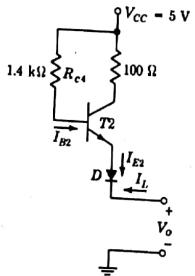
Providing as it does a mechanism for the rapid removal of the output transistor base charge, the TTL gate has only one principal remaining limitation on gate speed. This results from the effects of capacitance which appears across the output of the gate, from the collector of T_3 to ground. This capacitance is composed of the capacitance of the output transistor itself, of the capacitance to ground of wires which connect the gate output to other gates, and of the input capacitances of these other gates or other devices which are being driven. When T_3 is driven to cutoff, this output capacitance must charge from V_{cc} through the pull-up resistor R_c . If the output capacitance is C_o , the capacitance charged and the output rises from logic 0 to logic 1 with a time constant $R_c C_o$. This time constant can be reduced by reducing the resistance of R_c . Such a reduction, however, would increase the power dissipation in R_c and, of course, in transistor T_3 while T_3 is conducting. In addition, the reduction in R_c would make it more difficult to saturate T_3 .

The expedient used in TTL gates to hasten the charging of output capacitance substantially without introducing an unacceptable increase in power dissipation is shown in Fig. 6.4-1. Here, the pull-up resistor R_c in Fig. 6.3-2 is replaced with the active devices, transistor T_2 and diode D . This circuit is recognized to be an active pull-up similar to the active pull-up used in RTL gates (Fig. 4.7-1) as well as in DTL gates. We now discuss, qualitatively, the operation of the circuit of Fig. 6.4-1.

FIGURE 6.4-1
A TTL NAND gate.

It is intended, in the active pull-up circuit of Fig. 6.4-1, that when the output V_o is at logic 0, transistor T_3 will be in saturation and T_2 cut off. Alternatively, when the output goes from logic 0 to logic 1, T_3 is to cut off, T_2 is to go ON, and the output capacitance is to charge through the series combination of transistors T_3 and T_2 and diode D . This switching splitter used to provide the bases of T_3 and T_2 with voltages which swing in opposite directions so that when one transistor of the T_3 - T_2 totem-pole pair is being driven ON, the other is being driven OFF and vice-versa.

To appreciate the need for the diode D located between the output transistor pair, consider the situation when the inputs are all at logic 1, in which case the output should be at logic 0. With all inputs at logic 1, transistors T_4 and T_3 will both be in saturation (we shall verify later that such is indeed the case). Then the collector voltage of T_3 is $V_{ce}(sat) \approx 0.2$ V while the collector voltage of T_4 is $V_{ce4}(sat) + V_{be3} \approx 0.2 + 0.75 = 0.95$. In this case, if diode D were not present, the base-emitter voltage of T_2 would be $V_{c4} - V_{c3} = 0.95 - 0.2 = 0.75$ V, and T_2 would also be in saturation. We require however that at the logic 0 output T_2 be cut off. With the diode D present, the 0.75 V drop between the collectors of T_4 and T_3 must divide between diode D and the base-emitter junction of T_2 . In this case neither diode D nor transistor T_2 will be forward-biased sufficiently to pass any appreciable current.

FIGURE 6.5-1
A loaded TTL gate.FIGURE 6.5-2
Output circuit used to calculate V_o when T_3 and T_4 are cut off.

that T_2 is operating in its active region, the base current is $I_{B2} = I_{E2}/(h_{FE} + 1) = -I_L/(h_{FE} + 1)$. Equation (6.5-1) becomes, since $h_{FE} \approx h_{FE} + 1$,

$$V_o = V_{CC} + \frac{R_{ce4} I_L}{h_{FE}} - V_{BE2} - V_D \quad (6.5-2)$$

When T_3 is cut off, V_o is at its logic 1 level. To calculate the load current I_L operating in their inverse mode, just as for the input transistor in the basic transistors of which are shown in Fig. 6.5-1) transistors T_3 and T_4 are in saturation. Hence, $V_{B4} = V_{BE3} + V_{BE4} = 0.75 + 0.75 = 1.5$ V. The base current of T_4 is being supplied almost entirely through the collector junction of the input transistor. This input transistor is operating in its active (albeit inverse) mode, and so we shall allow a drop of 0.7 V across its base-collector junction. Altogether the voltage at the base of the input transistor is $1.5 + 0.7 = 2.2$ V. The drop across each base resistor R_b is therefore $5.0 - 2.2 = 2.8$ V. The current through R_b is $2.8/(4 \text{ k}\Omega) = 0.70 \text{ mA}$. Assuming that the input transistors $T_{11} - T_{1N}$ each have $h_{FC} = 0.02$, the input current to each driven gate is $0.70(0.02) = 14 \mu\text{A}$.

If the driving gate T_2 were fanned out to a single gate, with the emitter junction and diode carrying so small a current ($14 \mu\text{A}$) we could reasonably take $V_{BE2} = V_D = V_y = 0.65$ V. In this case also, the drop across R_{ce4} [the second term in Eq. (6.5-2)] is negligible, and we have

$$V_o \approx V_{CC} - 2V_y = 5.0 - 2(0.65) = 3.7 \text{ V} \quad (6.5-3)$$

Suppose, on the other hand, we drive 10 gates (as we shall see, a fan-out of 10 is the conservative figure generally specified by manufacturers) and suppose we allow as a worst case that h_{FC} may be as large as 0.1. In this case the load current would be of the order of 1 mA. Even with this current the drop across

6.5 INPUT-OUTPUT CHARACTERISTIC NEGLECTING THE INPUT TRANSISTOR

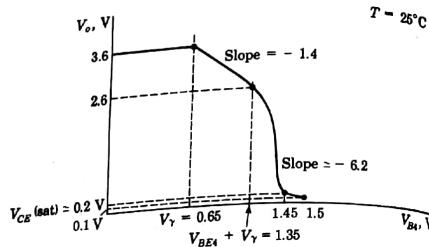
We shall now undertake to describe and make calculations concerning the operation of the TTL gate shown in Fig. 6.5-1. Since this gate is substantially more complicated than either the RTL gate or the DTL gate, we find it convenient to consider at the outset not the overall input-output characteristic but the characteristic relating the gate output voltage V_o to the base voltage V_{B4} of transistor T_4 . In Sec. 6.6 we shall consider the relation between V_{B4} and the input voltage V_i ; and by combining these two characteristics, we shall deduce the overall characteristic.

Let us start with $V_{B4} = 0$ V. In this case T_4 and T_3 are cut off. Transistor T_2 supplies a current I_{E2} to the N driven gates. The emitter current I_{E2} is equal to the current $-I_L$, I_L being the load current. Following common convention, the positive direction of I_L is taken into the gate.

The circuit from which to calculate the output voltage V_o is given in Fig. 6.5-2. The output voltage V_o is given by

$$V_o = V_{CC} - R_{ce4} I_{B2} - V_{BE2} - V_D \quad (6.5-1)$$

where V_{CC} (= 5 V) is the supply voltage, $R_{ce4} I_{B2}$ is the drop across R_{ce4} , V_{BE2} is the base-emitter voltage of T_2 , and V_D is the drop across the diode. Assuming

FIGURE 6.5-3
Input-output characteristic of the TTL gate.

R_{e4} would continue to be negligible, but it would be more reasonable to take $V_{BE2} = V_D = 0.75$ V. The output voltage would then be

$$V_o = V_{cc} - 2V_g = 5.0 - 2(0.7) = 3.6 \text{ V} \quad (6.5-4)$$

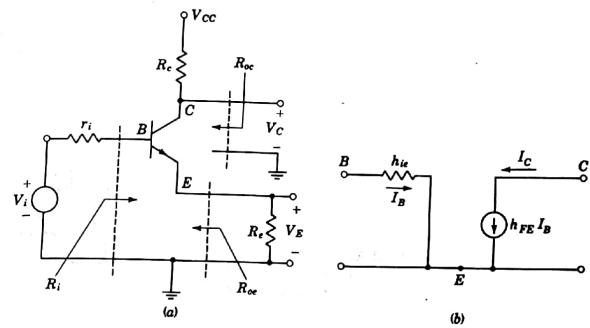
We therefore see that when V_o is in the 1 state, the value of V_o is only slightly dependent on the load current and hence on the fan-out. We shall accordingly simplify matters by assuming that when $T4$ (and hence $T3$) are cut off, the output voltage V_o is as given in Eq. (6.5-4). This value of the output voltage is given in the plot of the input-output characteristic shown in Fig. 6.5-3.

Phase splitter Before continuing, it will be well to digress briefly to review some of the gain and impedance characteristics of a phase splitter. Such a stage is shown in Fig. 6.5-4a. The transistor is assumed to be biased into its active region, although the biasing arrangements are not explicitly shown. The stage has a collector resistor R_c and an emitter resistor R_e and is driven by a source of impedance r_i . We assume that, in the active region, the transistor can be represented by the equivalent circuit involving hybrid parameters, as shown in Fig. 6.5-4b. We then find that the impedances and the voltage gains are given by the equations in Fig. 6.5-4. Rather typically we might assume that $h_{FE} = 50$ and $h_{ie} = 1 \text{ k}\Omega$. If $h_{FE} R_e \gg h_{ie} + r_i$, these equations for impedances and gains can be simplified to the useful approximations

$$R_i \approx h_{FE} R_e \quad (6.5-5a)$$

$$A_B = \frac{V_o}{V_i} \approx 1 \quad (6.5-5b)$$

$$A_C = \frac{V_C}{V_i} = -\frac{R_c}{R_e} \quad (6.5-5c)$$



$$R_i = h_{ie} + (h_{FE} + 1) R_e \quad A_B = \frac{V_o}{V_i} = 1 - \frac{h_{ie} + r_i}{R_i}$$

$$R_{oe} = \frac{h_{ie} + r_i}{h_{FE}} \quad A_C = \frac{V_C}{V_i} = -\frac{R_c}{R_e} A_B$$

$$R_{oe} = R_c$$

FIGURE 6.5-4
(a) A phase splitter. (b) Equivalent circuit of a transistor.

Returning to the circuit of Fig. 6.5-1, let us allow V_{B4} to increase from zero. When V_{B4} attains a value of about $V_{B4} = V_T = 0.65$ V, transistor $T4$ will begin to enter its active region. However, because of the drop across the base-emitter junction of $T4$, $T3$ will remain cut off. Hence, at this point, the circuit used to calculate the response of the output voltage V_o to V_{B4} is as shown in Fig. 6.5-5.

We now calculate the slope of the transfer characteristic in the region where $T3$ is cut off. The gain $A_2 \equiv \Delta V_o / \Delta V_{B2}$ from the base of $T2$ to the output is the gain of an emitter-follower and, hence, as in Eq. (6.5-5b) is $A_2 \approx 1$. The gain provided by $T4$ is [from Eq. (6.5-5c)] $A_4 = -R_{e4}/R_e = -1.4$. The overall gain $A = \Delta V_o / \Delta V_{B4}$ is $A_2 A_4 = -1.4$, as indicated in Fig. 6.5-3.

With further increase in V_{B4} , $T3$ will eventually reach the cut-in point. At this point $T4$ is in the active region, and we shall correspondingly assume a base-emitter voltage drop of $V_{BE4} = 0.7$ V. Assuming, as usual, that $T3$ comes out of cutoff when the voltage drop across its base-emitter junction is $V_{BE3} = V_T = 0.65$ V, we estimate that $T3$ begins to turn on when $V_{B4} = V_{BE4} + V_{BE3} = 0.7 + 0.65 = 1.35$ V. At the point where $T3$ just turns on, the current through R_{e4} and, hence, very nearly, the current through R_{e4} is $V_T/R_e = 0.65/1 \text{ k}\Omega = 0.65 \text{ mA}$. The corresponding drop through $R_{e4} = 1.4 \times 0.65 = 0.9$ V. Hence,

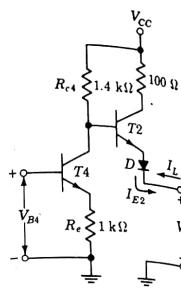


FIGURE 6.5-5
Circuit used to calculate V_o versus V_{B4}
when $T4$ is ON and $T3$ is still OFF.

as appears in Fig. 6.5-3, when $V_i = 1.35$ V, V_o is 0.9 V lower than the value in Eq. (6.5-4), that is $V_o = 3.5 - 0.9 = 2.6$ V.

When $T3$ turns ON, the incremental gain $A = \Delta V_o / \Delta V_{B4}$ increases and the output voltage V_o begins to drop more sharply with increasing V_{B4} than when $T3$ was OFF. This increase in gain has a twofold source. On the one hand V_o drops, simply because $T3$ begins to conduct. On the other, with $T3$ ON, the emitter resistor R_e is shunted by the impedance seen looking into the base of $T3$. The resistance in the emitter circuit of $T4$ is thereby decreased, and, as indicated in Eq. (6.5-5c), the gain of $T4$ increases. The following illustrative calculation will display these two sources of increased gain.

EXAMPLE 6.5-1 In the circuit of Fig. 6.5-1, let $T3$, $T2$, and $T4$ be assumed to be in the active region, each with parameters $h_{ie} = 1$ kΩ and $h_{FE} = 50$. Calculate the incremental gain $A = \Delta V_o / \Delta V_{B4}$.

SOLUTION With $h_{ie} = 1$ kΩ shunting $R_e = 1$ kΩ, the equivalent resistor in the emitter of $T4$ is $R'_e = 500$ Ω. The condition $h_{FE} R'_e \gg h_{ie}$ continues to apply. Hence, the gain from base to emitter of $T4$, which is also the gain from the base of $T4$ to the base of $T3$, continues to have the value unity. In Fig. 6.5-6 we have redrawn the relevant portion of Fig. 6.5-1, with R_e replaced by R'_e , the connection from the emitter of $T4$ to the base of $T3$ removed, and with a generator $\Delta V_{B3} = \Delta V_{B4}$ applied at the base of $T3$.

Let us now apply the principle of superposition and calculate separately the output voltage ΔV_o due individually to each of the generators ΔV_{B4} and ΔV_{B3} . The gain, referring to the generator ΔV_{B4} , is calculated as before, except that R_e is now replaced by R'_e . We then have

$$A_4 = \frac{\Delta V_{C4}}{\Delta V_{B4}} = -\frac{R_{c4}}{R'_e} = -\frac{1.4}{0.5} = -2.8 \quad (6.5-6)$$

Next we calculate the gain due to ΔV_{B3} . Referring to Fig. 6.5-4, we can readily verify

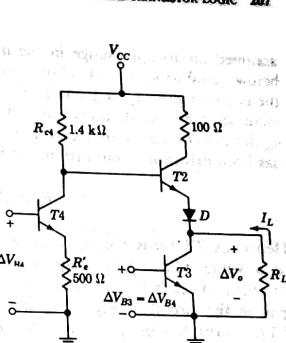


FIGURE 6.5-6
Circuit used to calculate incremental gain
when $T3$ is in the active region.

that a common-emitter amplifier stage operating without an emitter resistor and with a collector resistor R_c has a gain

$$A_c = \frac{\Delta V_o}{\Delta V_{B3}} = -\frac{h_{FE}}{h_{ie}} R_c \quad (6.5-7)$$

The equivalent collector resistor R_c with which $T3$ operates is the incremental impedance seen looking back into the diode D and the emitter of $T2$. The transistor impedance $\Delta V_{BE}/\Delta I_E \equiv h_{ie}$. The impedance $\Delta V_{BE}/\Delta I_E = \Delta V_{BE}/h_{FE} \Delta I_B = h_{ie}/h_{FE}$. Since the same current flows in the diode as in $T2$, we let the diode impedance equal the transistor impedance h_{ie}/h_{FE} . To find R_c we add the diode impedance to the impedance seen looking into the emitter of $T2$, as given by the equation for R_{oe} in Fig. 6.5-4. We have

$$R_c = \frac{h_{ie}}{h_{FE}} + \frac{h_{ie} + R_{c4}}{h_{ie}} = \frac{10^3}{50} + \frac{(1.4 + 1) \times 10^3}{50} = 68 \Omega \quad (6.5-8)$$

a value small enough in comparison with any load the gate will encounter to permit us to ignore the load. From Eqs. (6.5-7) and (6.5-8) we have

$$A_3 = \frac{\Delta V_o}{\Delta V_{B3}} \approx -\frac{h_{FE}}{h_{ie}} R_c = -\frac{1.4 + 2}{1} = -3.4 \quad (6.5-9)$$

The total incremental output voltage ΔV_o is therefore $\Delta V_o = A_4 \Delta V_{B4} + A_3 \Delta V_{B3}$. However, $\Delta V_{B3} = \Delta V_{B4} \approx \Delta V_{B4}$ since $T4$ is an emitter follower. Hence, $\Delta V_o = (A_4 + A_3) \Delta V_{B4}$ and the overall gain is

$$A = \frac{\Delta V_o}{\Delta V_{B4}} = A_4 + A_3 = -2.8 - 3.4 = -6.2 \quad (6.5-10)$$

The calculation in Example 6.5-1 indicates an abrupt change in the magnitude of the gain from 1.4 to 6.2 when $T3$ turns on. Actually we

assumed an abrupt change in the input impedance h_{ie} of $T3$ from h_{ie} infinite below cutoff to $h_{ie} = 1 \text{ k}\Omega$ above cutoff. It is the finite value of h_{ie} that yields the nonzero gain A_3 as given in Eq. (6.5-7), and it is the finite value of h_{ie} that increases the gain A_4 above its initial value of 1.4. However, the parameter h_{ie} depends on the emitter current of $T3$ and the extent to which the transistor has been driven into saturation. It is readily established that

$$h_{ie} \approx (1 + \sigma h_{FE}) \frac{V_T}{I_E} \quad (6.5-11)$$

Hence, as $T3$ turns ON, h_{ie} changes gradually from an extremely large value to a value which gets progressively smaller as $T3$ draws more and more current and tends toward saturation. Thus, the magnitude of the incremental gain in the circuit of Fig. 6.5-6 starts at 1.4 when $T3$ is OFF, increases when $T3$ turns ON, has the value 6.2 when $h_{ie} = 1 \text{ k}\Omega$, and increases still more as the current in $T3$ increases further. These features appear in the plot of the input-output characteristic shown in Fig. 6.5-3.

Referring once more to Fig. 6.5-1, we see that with still further increases of V_{B4} , $T3$ is driven eventually to saturation. Assuming, as we have, that at saturation the base-to-emitter voltage of a transistor is $V_o = 0.75 \text{ V}$, we would have saturation $V_o \approx 0.2 \text{ V}$ when

$$V_{B4} = V_{BE4} + V_o = 0.7 + 0.75 = 1.45 \text{ V} \quad (6.5-12)$$

This estimate has been included on the plot of Fig. 6.5-3.

When V_{B4} is increased beyond 1.45 V to 1.5 V both transistors $T4$ and $T3$ are saturated and $T2$ is cut off. In this case $T3$ is deeply saturated and $V_o \approx 0.1 \text{ V}$. Then collector current $I_{C3} = I_L$, and the load current is now positive. This current is calculated in Sec. 6.9.

6.6 INPUT-OUTPUT CHARACTERISTIC OF THE INPUT TRANSISTOR

The plot of Fig. 6.5-3 relates the output voltage V_o to the base voltage V_{B4} . Of much greater interest is the input-output characteristic that relates V_o to the input voltage V_i , shown in Fig. 6.6-1. We now study the operation of transistor $T1$ to enable us to determine the relation between V_i and V_o .

As we have noted, when V_i is at a high voltage corresponding to logic 1, the input transistor $T1$ operates in the active inverse mode. The collector current I_{C1} which is the base current I_{B4} of $T4$ is supplied principally by the base current I_{B1} of $T1$. The current gain h_{FC} in this inverse mode is extremely low by deliberate transistor design ($h_{FC} \approx 0.02$), so that the input source V_i supplies only about 2 percent of the base current of $T4$. Under these circumstances $T4$ and $T3$ are in saturation, $V_{B4} = 0.75 + 0.75 = 1.5 \text{ V}$, and the gate output V_o is at logic 0. This situation prevails so long as V_i is sufficiently positive with respect to V_{B4} ($2V_o = 1.5 \text{ V}$) to maintain $T1$ in the active inverse region.

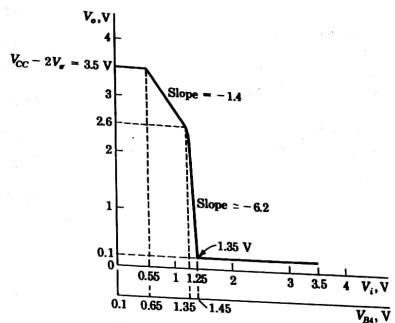


FIGURE 6.6-1
Input-output characteristic of a loaded TTL gate.

We shall now show that as V_i falls, then, at least up to the point where $V_i = 2V_o$ ($= 1.5 \text{ V}$), the transistors $T4$ and $T3$ remain in saturation and consequently V_o remains at logic 0. With $V_i = 2V_o$, $V_{CE1} = 0 \text{ V}$, and $T1$ is in saturation. From Eq. (1.10-11) we find that when $V_{CE1} = V_{CE}(\text{sat}) = 0$, $\sigma = -1/(h_{FE} + h_{FC}) \approx -1/h_{FE}$ since $h_{FE} \gg h_{FC}$. Applying the definition of σ to transistor $T1$, we have that $\sigma = I_{C1}/h_{FE} I_{B1}$. Substituting in this definition the value $\sigma = -1/h_{FE}$, we find that $-I_{C1} = I_{B1}$. Hence, the emitter current of $T1$ is now zero. However, the emitter current of $T1$ supplied only 2 percent of the base current of $T4$, and this small loss of base current will hardly affect $T4$ or $T3$. Hence, V_o remains at logic 0.

As V_i falls to voltages lower than $2V_o = 1.5 \text{ V}$, the base current I_{B1} will be diverted away from the base-collector junction of $T1$ into its base-emitter junction. Eventually, the base current of $T4$ will be reduced to a value where the transistor $T4$ will be at the verge of coming out of saturation. We calculate now the voltage V_i corresponding to this situation. We have often noted that when a transistor is comfortably inside the saturation region, its collector-emitter voltage is $V_{CE}(\text{sat}) = 0.2 \text{ V}$. At the very edge of saturation a more reasonable number is $V_{CE}(\text{sat}) = 0.3 \text{ V}$. Referring to Fig. 6.5-1, since $T2$ is OFF when $T3$ is ON, we have

$$V_{CC} = R_{ce} I_{C4} + V_{CE4}(\text{sat}) + V_{B3} \quad (6.6-1)$$

or

$$5 = 1.4 \times 10^3 I_{C4} + 0.3 + 0.75 \quad (6.6-2)$$

from which

$$I_{C4} = 2.8 \text{ mA} \quad (6.6-3)$$

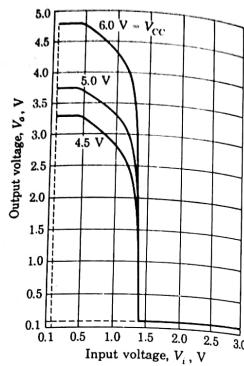


FIGURE 6.6-2
Medium-speed TTL transfer characteristics
 $V_{cc} = 4.5, 5.0$, and 6.0 V; $T_A = 25^\circ\text{C}$.

Since $T4$ is at the edge of saturation, $I_{C4} = h_{FE} I_{B4}$; therefore using $h_{FE} = 50$, we have

$$-I_{C1} = I_{B4} = \frac{I_{C4}}{h_{FE}} = \frac{2.8 \text{ mA}}{50} = 56 \mu\text{A} \quad (6.6-4)$$

The base voltage V_{B4} is 1.5 V. We allow a voltage drop of 0.7 V across the collector junction of $T1$ since this junction is carrying only $56 \mu\text{A}$. The base-to-ground voltage of $T1$ is $1.5 + 0.7 = 2.2$ V. Consequently the base current of $T1$ is $I_{B1} = (5 - 2.2)/(4 \text{ k}\Omega) = 0.7 \text{ mA}$. We can now calculate the parameter σ for transistor $T1$; we find

$$\sigma = \frac{I_{C1}}{h_{FE} I_{B1}} = \frac{-56 \times 10^{-6}}{50(0.7) \times 10^{-3}} = -1.6 \times 10^{-3} \quad (6.6-5)$$

This value of σ is so close to $\sigma = 0$ that we shall make no error in assuming $\sigma = 0$. In this case $V_{CE(\text{sat})}$ is given by Eq. (1.10-4), and we have

$$V_{CE1} \approx V_i \ln \frac{1}{\alpha_i} \quad (6.6-6)$$

For $\alpha_i = 0.02$, $V_{CE} \approx 100 \text{ mV}$. Thus, the voltage V_i at which $T4$ comes out of saturation is about 0.1 V lower than the corresponding voltage V_{B4} . Hence, while in Fig. 6.5-3, the saturation region begins at $V_{B4} = 1.5$ V, in Fig. 6.6-1 it begins at $V_i = 1.4$ V.

As V_i continues to decrease, the current $I_{B4} = -I_{C1}$ continues to decrease, becoming zero when $T4$ cuts off and remaining zero thereafter. Hence, σ , which has the value given by Eq. (6.6-5) when $T4$ is in saturation, falls to $\sigma = 0$ when

$T4$ cuts off and remains at zero thereafter. Hence, again V_i is 0.1 V lower than V_{B4} .

It is now apparent that a plot of the overall input-output characteristic (V_i versus V_o) of the TTL gate is the same as the plot of V_o versus V_{B4} shifted characteristic is shown in Fig. 6.6-1 and is to be compared with the characteristics shown in Fig. 6.6-2. These latter characteristics are typical measured characteristics supplied by manufacturers.

6.7 THE MULTIEmitter TRANSISTOR

In the previous section we ignored the fact that the input transistor has several emitters. We shall now take this feature into account. For convenience, in Fig. 6.7-1 we have represented the multiemitter transistor as an array of individual transistors with collectors joined and bases joined. The inputs $V_{iA}, V_{iB}, \dots, V_{iN}$ are applied to the individual emitters.

Suppose now that all inputs are at the logic level 1, corresponding nominally to $V_{iA} = V_{iB} = \dots = V_{iN} = V_{cc} - 2V_o = 3.5$ V. All input transistors $T1$ are now operating in the inverse active region. We assume tentatively that transistors $T3$ and $T4$ are in saturation. Thus, the voltage $V_{B4} = V_o + V_o = 1.5$ V while $V_{B1} = 0.7 + 1.5 = 2.2$ V. The current through R is then $I_{B1} = (5.0 - 2.2)/(4 \text{ k}\Omega) = 0.7 \text{ mA}$. In this mode of operation the combined emitter current $I_E = -h_{FC} I_{B1}$, where h_{FC} is the inverse current gain corresponding to α_i . Now $h_{FC} = \alpha_i/(1 - \alpha_i) \approx \alpha_i$ since $\alpha_i \ll 1$. Using $\alpha_i = 0.02$, we have $I_E = -0.02(0.7) \text{ mA} = -0.014 \text{ mA}$. Thus, the total base current of $T4$ is $I_{B4} \approx 0.7 \text{ mA}$. This is the same result calculated in Sec. 6.6 for a single-emitter input transistor $T1$.

We note again that the base current required to drive $T4$ comes almost entirely from V_{cc} through R . The total current required from the gate driving $T1A, T1B, \dots, T1N$ is 2 percent of the base current I_{B4} . If there are N inputs, each driving gate need supply only $2/N$ percent of this current.

Next, let us consider that one of the input emitters, say the emitter of $T1A$, is at logic level 0, which corresponds to about 0.2 V, the saturation voltage of the driving gate. Then, as we have seen in Sec. 6.6, $T1A$ will be in saturation, and $V_{B4} = V_{iA} + V_{CE(T1A)} \approx 0.2 + 0.1 = 0.3$ V. Transistors $T4$ and $T3$ are therefore cut off, as required. Now, while $T1A$ is operating in saturation with its emitter junction forward-biased and carrying an emitter current very large in comparison with its collector current, let us assume that the emitters of each of the other "transistors" $T1B, \dots, T1N$ are in the logic 1 state, with $V_i = 3.5$ V. These transistors are operating in the inverse active mode with their base-collector junctions forward-biased and their emitters reverse-biased. We note further that since $V_{iA} = 0.2$ V and $T1A$ is saturated, $V_{B1} = V_{iA} + V_{BE(T1A)} = 0.2 + 0.75 = 0.95$ V. Also, the collector voltages of transistors $T1B, \dots, T1N$ are equal to V_{B4} , where $V_{B4} = 0.3$ V. Hence, the voltage drops across the

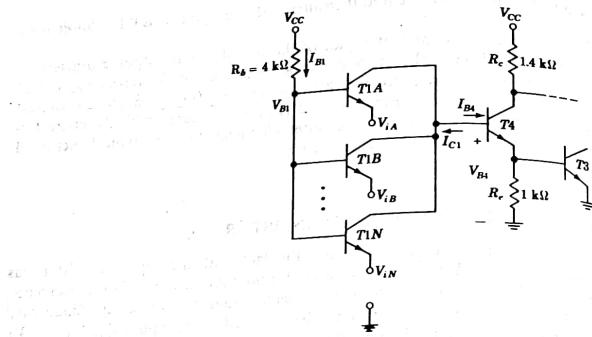


FIGURE 6.7-1
Circuit to show operation of the multiemitter transistor.

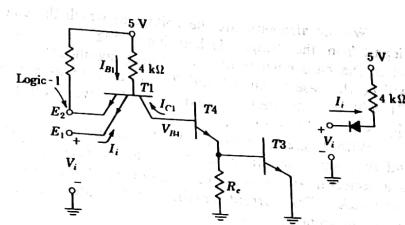
forward-biased base-collector junctions of $T1B, \dots, T1N$ are 0.65 V, which is 0.1 V less than the forward bias across the base-emitter junction of $T1A$. On many occasions, we have noted that a change of 0.1 V in voltage will carry a transistor from saturation to the edge of cutoff. Hence, in Fig. 6.7-1, we may estimate that when one (or more) inputs are set at the logic level 0, the other transistors are, to all intents and purposes, cut off.

One additional characteristic of the multiemitter transistor should be mentioned. Referring to Fig. 6.3-1, we see that transistors are formed not only between each of the three emitters and the collector but also between the emitters E_A and E_B , E_A and E_C , and E_B and E_C since they also are *npn* structures. However, because of the geometry of the "transistors," their current gains are very small (less than 0.01) and the presence of these transistors may be neglected.

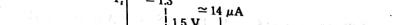
6.8 INPUT VOLT-AMPERE CHARACTERISTIC OF THE TTL GATE

Because of the relative complexity of TTL logic, it is useful to consider the input and the output volt-ampere characteristic in addition to the transfer characteristic. We consider here the input characteristic. The output characteristic is discussed in Sec. 6.9.

In Fig. 6.8-1a we are interested in the input volt-ampere characteristic at one emitter input. Other emitter inputs (only one additional input is shown in the figure) are assumed to be at logic level 1. For the reasons presented in Sec. 6.7 we neglect the effect of these other emitter terminals.



(a)



(b)

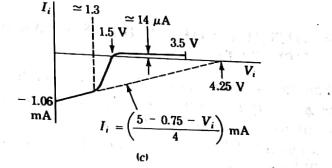


FIGURE 6.8-1
Input volt-ampere characteristic.

When $V_i = 0$, $T4$ is cut off and therefore the collector current $I_{C1} = 0$. In this case, neglecting the current contribution from E_2 (see Sec. 6.7), we can represent $T1$ as a diode. An equivalent circuit for calculating the current I_i is then shown in Fig. 6.8-1b. When we assume the usual 0.75 V junction voltage, $I_i = -(5 - 0.75)/(4 \text{ k}\Omega) = -1.06 \text{ mA}$, as noted in the plot of Fig. 6.8-1c. As long as all the current through the 4-kΩ resistor flows through the E_1 junction, the volt-ampere characteristic is a straight line with current intercept -1.06 mA and voltage intercept $5 - 0.75 = 4.25 \text{ V}$. This straight line is indicated (dashed) in the plot.

As V_i increases, the V_i - I_i plot departs from the straight line as $T4$ begins to turn on. When $I_i = 0$, all the current through the 4-kΩ resistor flows into the collector of $T1$ and into the base of $T4$. At this point both $T4$ and $T3$ will be in saturation, and the voltage V_{B4} will be 1.5 V. Since $I_i = 0$, the emitter current in $T1$ is zero and the collector-emitter voltage is given by Eq. (1.10-9):

$$V_{CE1} = -V_T \ln \frac{1}{\alpha_N} \quad (6.8-1)$$

With $V_T = 25 \text{ mV}$ and $\alpha_N = 0.98$, we have $V_{CE1} \approx -0.5 \text{ mV}$. Thus, the input voltage V_i is 0.5 mV less than $V_{B4} = 1.5 \text{ V}$. Hence, we can approximate $V_i \approx 1.5 \text{ V}$ when $I_i = 0$. This point $V_i = 1.5 \text{ V}$ and $I_i = 0$ is noted in the plot of Fig. 6.8-1c.

We can also estimate the voltage at which the characteristic will begin to depart from the straight dashed line in Fig. 6.8-1c. This departure will begin when the base current of T_4 becomes significant. We may judge that such will be the case when T_4 and T_3 are each in its active region. At this time T_1 will be saturated, so that altogether we shall have $V_i = V_{BE3} + V_{BE4} \sim V_{CE1} \approx 0.7 + 0.7 - 0.1 = 1.3$ V.

With V_i greater than 1.5 V, the emitter current in T_1 reverses direction and the transistor enters its inverse active region. The emitter E_1 then acts as a collector and $I_i = h_{FE} I_{B1}$. Using $h_{FE} \approx 0.02$ and $I_{B1} = 0.7$ mA, we have $I_i \approx 14 \mu\text{A}$. This current remains approximately constant as V_i increases to its maximum voltage.

6.9 OUTPUT VOLT-AMPERE CHARACTERISTIC OF THE TTL GATE

When the gate input in Fig. 6.5-1 is at logic 0, transistors T_4 and T_3 are cut off. The gate output is now at logic 1, and the gate is *sourcing* current, i.e., furnishing current to a load so that I_L is negative. When the gate input is at logic 1, T_2 is cut off, T_3 is in saturation, and the gate will be *sinking* current, that is, I_L will be positive.

Consider first the case where the input is at logic level 0, T_3 is cut off, and the gate is sourcing current. Then the output characteristic of the gate can be deduced from Fig. 6.9-1, where $-I_L = I_S$ is the source current. Since, in this mode, transistor T_2 is in the active region, the base current is $I_S/(h_{FE} + 1)$. Starting at the 5 V supply and taking account of the voltage drops across R_{c4} across the base-emitter junction of T_2 (0.65 V), and across the diode D (0.65 V), we find

$$V_o = 3.7 - \frac{R_{c4}}{h_{FE} + 1} I_S \quad (6.9-1)$$

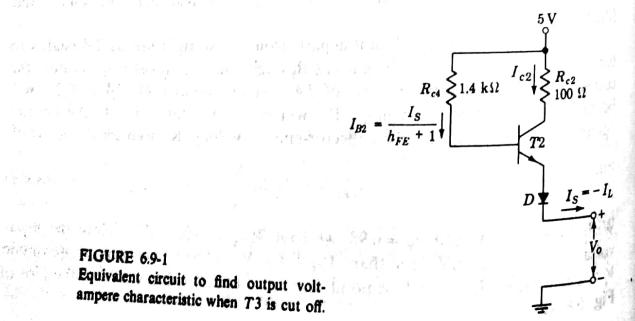


FIGURE 6.9-1
Equivalent circuit to find output volt-ampere characteristic when T_3 is cut off.

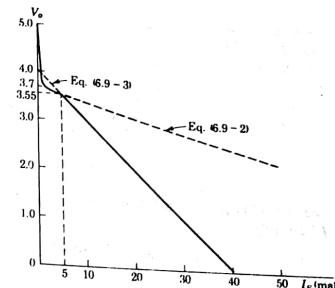


FIGURE 6.9-2
Output-input characteristics when T_3 is cut off.

Thus, the output has a Thevenin representation which consists of a voltage of 3.7 V in series with an output impedance $R_{c4}/(h_{FE} + 1)$. With $R_{c4} = 1.4 \text{ k}\Omega$ and using $h_{FE} + 1 = 50$, we find that the output impedance is 28Ω , so that

$$V_o = 3.7 - 28 I_S \quad (6.9-2)$$

When transistor T_2 is in saturation, $V_{BE2} = 0.75$ and $V_{CE2} = 0.2$ V. Also $V_D = 0.75$ V. In this case $I_{B2} = (3.5 - V_o)/1,400$ and $I_{C2} = (4.05 - V_o)/100$. Since $I_S = I_{C2} + I_{B2}$ and $I_{B2} \ll I_{C2}$, we have

$$V_o \approx 4.05 - 100 I_S \quad (6.9-3)$$

The straight-line plots for Eqs. (6.9-2) and (6.9-3) intersect at $I_S \approx 5$ mA. Hence, for $I_S > 5$ mA the transistor is in saturation, and Eq. (6.9-3) applies, while for $I_S < 5$ mA the transistor is in the active region, and Eq. (6.9-2) applies. The output characteristic is then given by the solid-line plot of Fig. 6.9-2. At very low currents that plot must depart from the straight line of Eq. (6.9-2). For at $I_S = 0$, there need be no drop across R_{c2} nor any drop across the base-emitter junction of T_2 or the diode D , and, in principle, V_o should become equal to $V_{cc} = 5.0$ V.

When all the inputs to the gate of Fig. 6.5-1 are at logic level 1, T_2 is cut off while T_3 and T_4 are in saturation. The output looks back directly across the saturated transistor T_3 , as in Fig. 6.9-3, which now sinks a current I_L . The volt-ampere characteristic at these output terminals is now precisely the common-emitter collector characteristic of the transistor corresponding to the base current I_{B3} . This characteristic is similar to that shown in Fig. 1.10-1, the difference being that we now plot $V_o = V_{CE3}(\text{sat})$ as a function of $I_L = \sigma h_{FE} I_{B3}$ rather than as a function of σ . We calculate I_{B3} as follows. Referring to Fig. 6.5-1, we have

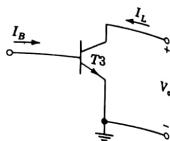


FIGURE 6.9-3
Equivalent circuit for calculating output volt-ampere characteristic when T_3 is saturated (T_2 is cut off).

(see Sec. 6.5) $I_{B1} \approx 0.7 \text{ mA}$ and therefore $I_{B4} \approx 0.7 \text{ mA}$. With T_4 saturated and T_2 cut off,

$$I_{C4} = \frac{V_{CC} - V_{BE3} - V_{CE4(\text{sat})}}{R_{c4}} = \frac{5 - (0.75 + 0.2)}{1.4 \text{ k}\Omega} \approx 2.9 \text{ mA} \quad (6.9-4)$$

Hence $I_{E4} = I_{B4} + I_{C4} = 0.7 + 2.9 = 3.6 \text{ mA}$

$$(6.9-5)$$

Since the current in resistor R_e is

$$I_{R_e} = \frac{0.75}{1 \text{ k}\Omega} = 0.75 \text{ mA} \quad (6.9-6)$$

we have

$$I_{B3} = I_{E4} - I_{R_e} = 3.6 - 0.75 = 2.85 \text{ mA} \quad (6.9-7)$$

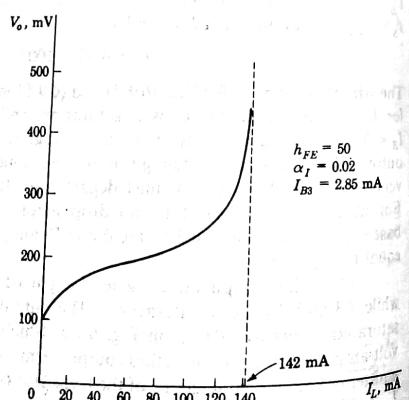


FIGURE 6.9-4
Output volt-ampere characteristic when all inputs to a TTL gate are in state 1.

A typical TTL output characteristic is shown in Fig. 6.9-4. This characteristic corresponds to a base current $I_{B3} = 2.85 \text{ mA}$ as given by Eq. (6.9-7). At low currents the output is $V_{ce}(\text{sat}) \approx 0.1 \text{ V}$, the value we have associated with an unloaded transistor in saturation. At higher currents the drop across the load $I_L = h_{FE} I_{B3} = 50(2.85 \times 10^{-3}) = 142 \text{ mA}$, transistor T_2 comes out of saturation.

6.10 MANUFACTURER'S DATA AND SPECIFICATIONS; TEMPERATURE DEPENDENCE AND NOISE IMMUNITY

As with other gates, the characteristics of TTL gates are temperature-dependent. As before, the principal source of this dependence is the temperature dependence of the base-emitter and base-collector junction voltages. Because of the similarity with calculations already made, these calculations of temperature effects are left for the problems. Instead we shall take note of this temperature dependence as it is evidenced in typical average characteristics published by manufacturers.

Figures 6.10-1 to 6.10-4 all apply to the Texas Instrument type 54/74 gate. The type 74 gates are intended for the ambient temperature range 0 to 70°C and allow for a supply-voltage range from 4.75 to 5.25 V. The type 54 are held to tighter tolerances, are intended for a temperature range -55 to 125°C, and allow a supply range from -4.5 to 5.5 V. This gate is a "standard" type TTL NAND gate and represents a useful compromise between good speed and modest power dissipation. Figure 6.10-1 gives typical transfer characteristics for a variety

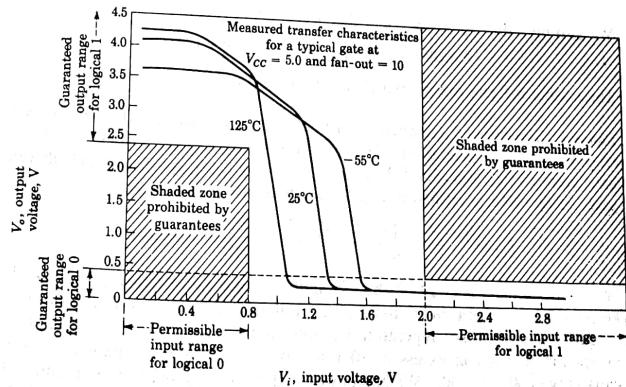


FIGURE 6.10-1
Voltage transfer characteristics for typical SNS54/74 TTL NAND gate.

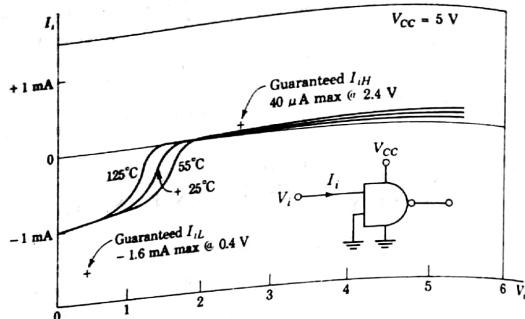


FIGURE 6.10-2
Input characteristics as a function of temperature for standard series SNS475.

of temperatures and for a fan-out of 10. Note that the manufacturer guarantees that any input voltage greater than 2.0 V will be acknowledged by the gate to correspond to the logic level 1; that is, $V_{iH} = 2.0$. Similarly it is guaranteed that the gate output at logic level 1 will never be less than 2.4 V; that is, $V_{oH} = 2.4$ V. Thus, the $\Delta 1$ noise margin is

$$\Delta 1 = V_{oH} - V_{iH} = 2.4 - 2.0 = 0.4 \text{ V}$$

We also note that $V_{iL} = 0.8$ V and that $V_{oL} = 0.4$ V, yielding

$$\Delta 0 = V_{iL} - V_{oL} = 0.8 - 0.4 = 0.4 \text{ V}$$

These noise margins are, as expected, extremely conservative.

Fan-out The fan-out is limited by the amount of current $T3$ (Fig. 6.5-1) can sink when it is in saturation. We find from Fig. 6.10-2 that when a driving gate output is at logic level 0, it must sink about 1.0 mA from each driven gate. This result is verified in Fig. 6.5-1, where, with $T3$ saturated,

$$I_{11} = \frac{V_{CC} - V_{BE}(T11) - V_o}{R_b} = \frac{5 - 0.75 - 0.2}{4 \text{ k}\Omega} \approx 1 \text{ mA}$$

We note also from Fig. 6.10-3 that the ability of the gate to sink current while keeping $T3$ in saturation is most severely limited at the lowest temperature, -55°C . Even at this temperature we note that $T3$ does not leave saturation except for currents in excess of 30 mA. We might then estimate a fan-out capability of $30/1.0 = 30$. We find however, that the manufacturer recommends a fan-out of only 10 to keep V_{oL} well below 0.4 V. Further the fan-out affects

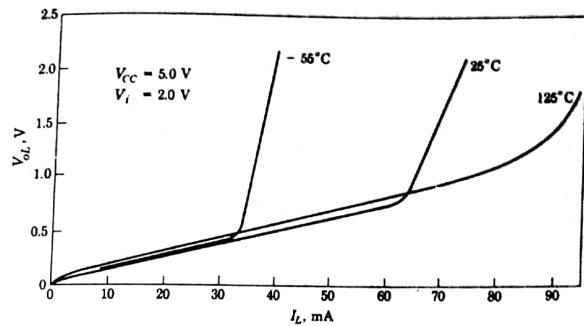


FIGURE 6.10-3
Output volt-ampere characteristic when all inputs are in 1-state (T3 saturated).

Figure 6.10-4 is a plot of the output voltage V_{oH} when the gate is in the high state, as a function of the sourcing current, I_S . Note that if $V_{oH} > 2.5$ V $I_S < 11$ mA at room temperature.

Propagation delay time Propagation delays in TTL gates are defined and measured in much the same manner as with other gates. The times $t_{pd}(HL)$ and $t_{pd}(LH)$ are as defined in Fig. 6.10-5a. Here we see that the times are

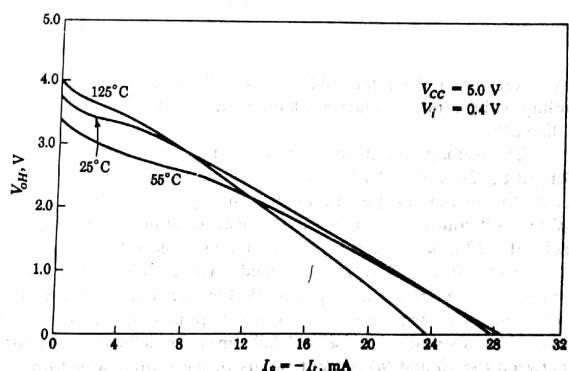


FIGURE 6.10-4
Output volt-ampere characteristic when V_i is in the 0 state (T3 cut off).

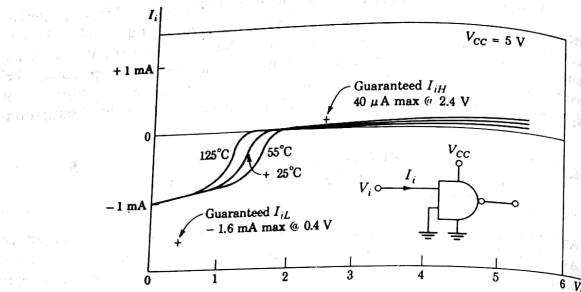


FIGURE 6.10-2
Input characteristics as a function of temperature for standard series SN5475.

of temperatures and for a fan-out of 10. Note that the manufacturer guarantees that any input voltage greater than 2.0 V will be acknowledged by the gate to correspond to the logic level 1; that is, $V_{IH} = 2.0$. Similarly it is guaranteed that the gate output at logic level 1 will never be less than 2.4 V; that is, $V_{OH} = 2.4$ V. Thus, the $\Delta 1$ noise margin is

$$\Delta 1 = V_{OH} - V_{IH} = 2.4 - 2.0 = 0.4 \text{ V} \quad (6.10-1)$$

We also note that $V_{IL} = 0.8$ V and that $V_{OL} = 0.4$ V, yielding

$$\Delta 0 = V_{IL} - V_{OL} = 0.8 - 0.4 = 0.4 \text{ V} \quad (6.10-2)$$

These noise margins are, as expected, extremely conservative.

Fan-out The fan-out is limited by the amount of current $T3$ (Fig. 6.5-1) can sink when it is in saturation. We find from Fig. 6.10-2 that when a driving-gate output is at logic level 0, it must sink about 1.0 mA from each driven gate. This result is verified in Fig. 6.5-1, where, with $T3$ saturated,

$$I_{11} = \frac{V_{CC} - V_{BE(T11)} - V_o}{R_b} = \frac{5 - 0.75 - 0.2}{4 \text{ k}\Omega} \approx 1 \text{ mA}$$

We note also from Fig. 6.10-3 that the ability of the gate to sink current while keeping $T3$ in saturation is most severely limited at the lowest temperature, -55°C . Even at this temperature we note that $T3$ does not leave saturation except for currents in excess of 30 mA. We might then estimate a fan-out capability of $30/1.0 = 30$. We find however, that the manufacturer recommends a fan-out of only 10 to keep V_{OL} well below 0.4 V. Further the fan-out affects propagation delay time as well as saturation.

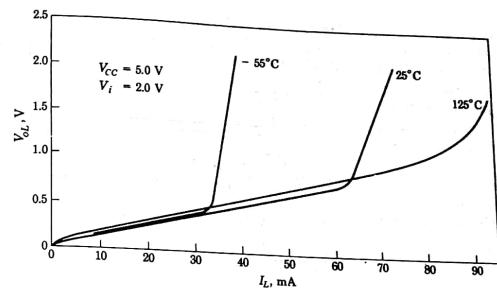


FIGURE 6.10-3
Output volt-ampere characteristic when all inputs are in 1-state (T3 saturated).

Figure 6.10-4 is a plot of the output voltage V_{OH} when the gate is in the high state, as a function of the sourcing current, I_S . Note that if $V_{OH} > 2.5$ V $I_S < 11$ mA at room temperature.

Propagation delay time Propagation delays in TTL gates are defined and measured in much the same manner as with other gates. The times $t_{pd}(HL)$ and $t_{pd}(LH)$ are as defined in Fig. 6.10-5a. Here we see that the times are

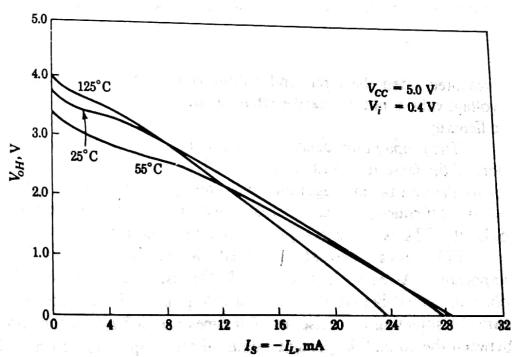


FIGURE 6.10-4
Output volt-ampere characteristic when V_i is in the 0 state (T3 cut off).

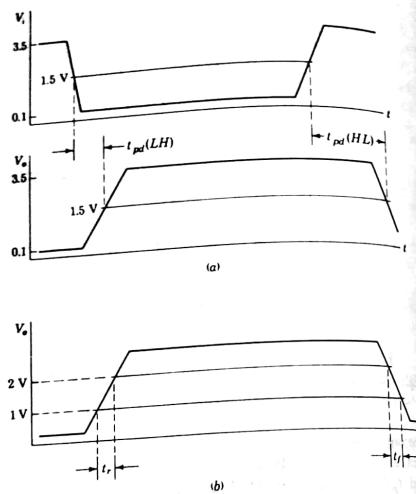


FIGURE 6.10-5
(a) Propagation delay time and (b) rise and fall time in TTL

measured when the input and output voltages are each equal to 1.5 V. The voltage value is not standard, each manufacturer defining the delay time somewhat differently.

The propagation delay times are a function of capacitive loading and therefore of the fan-out. With a capacitive load of the order of 15 pF and a fan-out of 10, the average propagation delay time $t_{pd} \equiv \frac{1}{2}[t_{pd}(HL) + t_{pd}(LH)]$ is typically 10 ns. Of course, increased capacitive loading increases the propagation delay. Schottky TTL (see Sec. 6.13) gates have propagation delay times of 2 to 4 ns.

TTL gates are used at high speeds, where the *rise time* and the *fall time* are important. As noted in Fig. 6.10-5b, the rise and fall times of TTL gates are often measured between the 1- and 2-V points of the output voltage. Alternatively, sometimes the rise and fall times are defined as the time of transition between the 10 and 90 percent points of the output waveform. Typical values are $t_r = 8$ ns and $t_f = 5$ ns for a medium-speed TTL gate and 4 to 8 ns for a high-speed TTL gate. Schottky TTL gates have rise and fall times of about 3 ns.

Because of its active pull-up, the TTL, like the RTL buffer (see Sec. 4.7), can accommodate a heavy capacitive load like that presented by 10 driven gates. When V_i drops to logic 0, T_4 cuts off and the base voltage of T_2 rises abruptly. The capacitive load due to the fan-out keeps the output voltage V_o initially at 0.1 V. Hence, as is easily verified, T_2 saturates and the current initially flowing out of the emitter of T_2 , $I_{E2} \approx 38.5$ mA. This large transient current (current spike) rapidly changes the load capacitance bringing V_o abruptly to the logic 1 level.

Unfortunately, it turns out that, when T_2 turns ON and saturates, it will do so before T_3 comes out of saturation. As a consequence a substantial portion of the current I_{E2} which should be available to charge the loading capacitance will instead be diverted into T_3 .

6.11 POWER-SUPPLY CURRENT DRAIN

The current drain from the power supply used with TTL gates is not the same in the two logic levels. The drain is larger for logic level 0 than for logic level 1. In a typical case, say in the gate of Fig. 6.5-1, the steady-state drain is roughly 3 mA in one level and roughly 1 mA in the other (see Prob. 6.11-1). The difference depends principally on the fact that in one case T_4 conducts and in the other is cut off. Of much more importance, however, is the fact that the current spike in I_{E2} , discussed in the preceding section, must be supplied by the power supply.

The exact waveform of the supply current depends both on the type of TTL gate involved and on the capacitive loading. Spikes range in amplitude up to about 40 mA, as noted, and have durations from several nanoseconds to many tens of nanoseconds in the case of heavy capacitive loading. A typical current waveform is shown in Fig. 6.11-1.

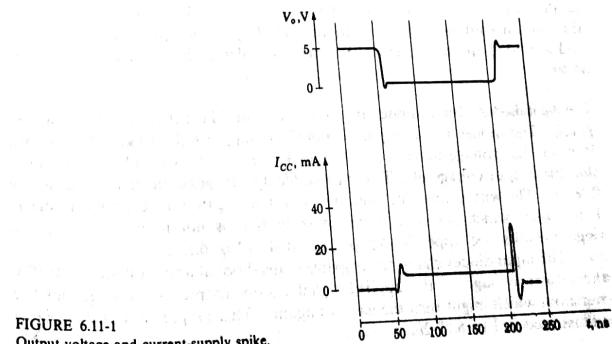


FIGURE 6.11-1
Output-voltage and current-supply spike.

These current spikes occur as frequently as the gate makes transitions, and hence the average current drawn is a function of the frequency at which the gate is being operated. It is not unusual for the average current drain to increase by a factor of 2 or 3 when the gate is being operated at some tens of megahertz.

If permitted to flow through the internal impedance of the power supply, these spikes of current would produce spikes of voltage, which would then be distributed throughout the system being serviced by the power supply. To avoid such a situation, it is necessary to provide capacitive bypassing to ground at the point where the power-supply lead is connected to the integrated-circuit chip.

6.12 TYPES OF TTL GATES

A number of types of TTL gates are available, differing principally in the compromise made between speed and power dissipation. Thus, the gate of Fig. 6.5-1 is considered a medium-speed gate and has an average power dissipation of about 10 mW and, as noted, a propagation delay of about 10 ns. The power dissipation can be decreased, at the expense of speed, by increasing the resistor values. Thus, we find that a commercially available gate with (see Fig. 6.5-1) $R_s = 40 \text{ k}\Omega$, $R_{e1} = 20 \text{ k}\Omega$, $R_e = 12 \text{ k}\Omega$, and $R_{e2} = 500 \Omega$ has a power dissipation of only 1 mW, but its average propagation time is 33 ns.

The speed of a gate can be increased by lowering resistor values. However, when speed is at a premium, additional changes are incorporated into the gate. A typical high-speed TTL gate is shown in Fig. 6.12-1. We note that R_s has been reduced to 2.4 k Ω , R_{e1} to 800 Ω , and R_{e2} to 60 Ω . We note three additional changes as well: (1) Diodes shunt each input to ground; (2) the emitter resistor of T_4 has been replaced by a circuit consisting of an additional transistor T_5 and two resistors; and (3) the connection from the collector of T_4 to the base of T_2 is no longer direct, as in Fig. 6.5-1, but is made instead through an emitter follower, forming a Darlington amplifier, composed of T_6 and a 3.5-k Ω emitter resistor. We now consider each of these modifications in turn.

The input diodes Input diodes are common to all TTL gates, except the slowest gates. The diodes act as input "clamps" to suppress the ringing that results from the fast voltage transitions found in TTL systems. Consider, for example, that the output voltage of a TTL gate suddenly changes from the 1 level to the 0 level. The wire connecting this gate to a driven gate now carries this signal. If the wire, which acts like a transmission line, is not terminated properly, ringing results (see Appendix A), as illustrated in Fig. 6.12-2a.

The input diodes clamp the negative undershoot at approximately -0.75 V and absorb enough of the applied signal energy to prevent a large positive overshoot which might turn the gate on again. This suppression of the ringing is illustrated in Fig. 6.12-2b.

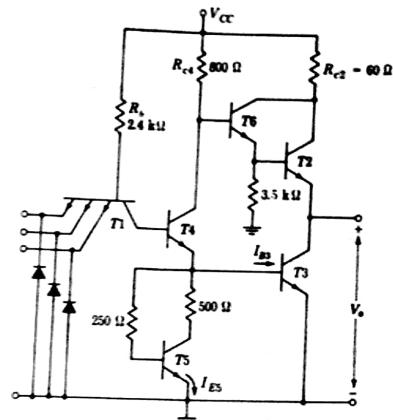


FIGURE 6.12-1
A high-speed TTL gate.

The Darlington circuit We consider now the effect of T_6 in the circuit of Fig. 6.12-1. First, we may note, as a matter of incidental interest, that the diode D encountered in the circuit of Fig. 6.5-1 is not used in Fig. 6.12-1. It will be recalled that the diode was included to ensure that T_2 will be cut off when T_4 and T_3 are saturated. In the configuration shown here the voltage drop across the base-emitter junction of T_6 serves the same function provided by the diode voltage drop, and T_2 remains cut off when T_3 is on.

Returning for a moment to the circuit of Fig. 6.5-1, we recall that when T_2 is in its active region, the output resistance seen at the gate output terminal is approximately R_{e1}/h_{FE} , for which, in Sec. 6.9 [see Eq. (6.9-2)] we estimated the value 28 Ω . In Fig. 6.12-1, we would calculate the output resistance as follows. First the output resistance, seen looking back into the emitter of T_6 , is $R_{e3}/h_{FE} = 800/50 = 16 \Omega$. Then, repeating the calculation for T_2 , we find that the gate output resistance is $16/h_{FE} = 16/50 \approx 0.3 \Omega$. We have neglected the resistance of the transistor in these calculations. As a result, we find that in the gate of Fig. 6.12-1 the measured output resistance is more nearly 10 Ω . In any event, the important characteristic of the gate of Fig. 6.12-1, is that when T_6 and T_2 are conducting, the gate output resistance is substantially lower than in the circuit of Fig. 6.5-1. (We recognize, of course, that transistors T_6 and T_2 are in a Darlington configuration, one characteristic of which is precisely this reduced output resistance.) This lowered output impedance increases

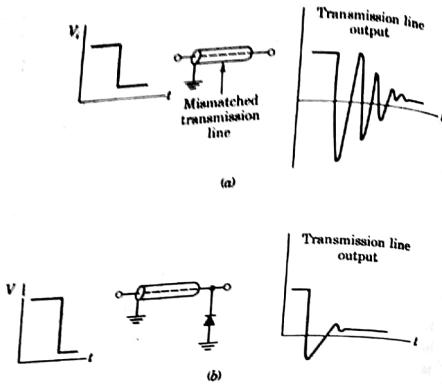


FIGURE 6.12-2
(a) Output of a transmission line to a negative-transition step, showing ringing.
(b) Output of the transmission line loaded by a diode.

the speed of operation of the gates. For with the lowered output resistance, any capacitance shunting the gate output will be able to charge more rapidly.

Of course, these considerations concerning output resistance do not apply when transistors T_6 and T_2 are cut off or in saturation. However, over a considerable range of the transition region from one logic level to the other, both transistors are in the active region, and the above discussion does apply. In this connection it is of interest to note that while in Fig. 6.5-1 transistor T_2 goes to saturation when the gate output goes to logic level 1, such is not the case in the circuit of Fig. 6.12-1. In a Darlington circuit only the input transistor (T_6 in Fig. 6.12-1) and not the output transistor (T_2) can be driven to saturation. For no matter whether T_6 is saturated or in its active region, V_{CE6} is positive. Since $V_{CE2} = V_{CE6}$, the collector-base junction of T_2 can never be forward-biased. Hence T_2 can never saturate.

The active pull-down The emitter resistor R_e in Fig. 6.5-1 provides the connection to ground for the base of the output transistor T_3 . Thus, R_e is the *pull-down* resistor which pulls the base of T_3 down to ground when T_4 cuts off. This terminology is analogous to use of the term *pull-up* resistor used in connection with collector resistors which are returned to the supply voltage. Extending the analogy, we shall refer to the circuit associated with T_5 in Fig. 6.12-1 as an *active pull-down*. We consider now the advantages which accrue from the use of such an active pull-down.

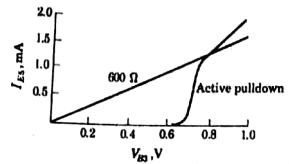


FIGURE 6.12-3
The effect of the active pull-down.

In Fig. 6.12-3, we compare the volt-ampere characteristics of a 600Ω resistor and the active pull-down. Calculations to show that the curve representing the active volt-ampere pull-down characteristic is reasonable are left as a student exercise. In high-speed TTL gates which use the Darlington configuration (T_6 and T_2 in Fig. 6.12-1) but which use a passive pull-down, the pull-down resistor has the value 600Ω ; hence the selection of a 600Ω resistor for comparison.

Now consider that transistor T_3 is to be turned ON. It will go ON when its base voltage exceeds 0.65 V and will reach saturation when the base voltage is 0.75 V . We note from Figs. 6.12-1 and 6.12-3 that over this range of voltages the active pull-down diverts less current from the base of T_3 than it would with a passive 600Ω pull-down resistor. Hence, in this mode of operation T_5 appears as a "resistor" which is larger than 600Ω . As a result we have as a first advantage that T_3 turns on faster.

To see a second advantage we note that when T_3 is ON and the fan-out is small, the base current in T_3 is substantially larger than is required to bring it to saturation. This excessive base current prolongs the storage time and delays the turnoff of the transistor. It also develops that T_3 can be driven so far into saturation that its base-emitter voltage may well exceed even 0.8 V . In such a case, again to be noted in Fig. 6.12-3, the active pull-down diverts more current from the base than a passive pull-down would. Hence, in this mode the active pull-down appears as a smaller resistor than the 600Ω resistance. Thus, excessive base current is somewhat suppressed.

In considering the *turnoff* of T_3 , we might imagine, however, that once the base voltage drops below 0.8 V , the active pull-down would be at a disadvantage, since it draws less current than a 600Ω resistor out of the base of T_3 . Such, however, is not the case. We must keep in mind that the active pull-down characteristic plotted in Fig. 6.12-3 applies only in steady state. Actually, at turnoff, the active pull-down continues to draw the larger current for some few nanoseconds until the charge distribution within T_5 itself has adjusted to the change.

The active pull-down also offers an important advantage in connection with the operation over a wide range of temperature. In a TTL gate without active pull-down, the turnoff time for T_3 increases with increasing temperature because T_3 is driven further into saturation. The reasons for the increases are twofold:

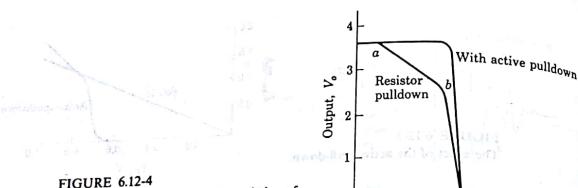


FIGURE 6.12-4
Comparison of transfer characteristics of TTL gates: resistor pull-down versus active pull-down.

(1) since junction voltages decrease with increasing temperature, with an increase in temperature more current will flow into the base of T_3 ; (2) the current gain of a transistor increases with temperature. Hence, even a fixed base current will drive a transistor farther into saturation as the temperature increases. With higher temperature, however, the active pull-down shown in Fig. 6.12-1 actually draws more current, thereby keeping I_{B3} fairly constant. This compensation results since an increase in temperature decreases V_{BE3} , thereby increasing I_{B3} and I_{C3} . A calculation to determine the change in I_{B3} resulting from a change in temperature is left as a student exercise.

A final advantage of the active pull-down is to be seen in its effect on the input-output voltage characteristic of the TTL gate. Characteristics with active pull-down and resistor pull-down are compared in Fig. 6.12-4. With resistor pull-down the plot exhibits a region (between a and b), as discussed in Sec. 6.5, where the slope has a magnitude equal to the ratio of collector to emitter resistors of T_4 . In this region T_4 is in its active region and provides gain. Point a marks the turn-on of T_4 , and point b marks the turn-on of transistor T_3 .

In the circuit of Fig. 6.12-1, until such time as T_3 turns on, the active pull-down provides no path for the emitter current of T_4 . Thus, T_4 and T_3 go on simultaneously, and the region $a-b$ is absent, with active pull-down. As a result the characteristic exhibits the much more abrupt change between levels, as indicated. The fact that the transition between logic levels is achieved with much smaller change in input voltage is, of course, of advantage in the matter of noise immunity.

6.13 SCHOTTKY TTL

In all TTL gates, transistors are driven into saturation and diodes are turned ON. A good part of the speed limitation of the gates results from the storage time delay associated with turning OFF and allowing transistors to recover from

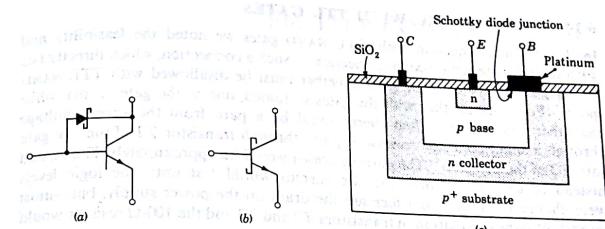


FIGURE 6.13-1
(a) Schottky diode connected from base to collector. (b) Schottky transistor. (c) Fabrication of the Schottky transistor.

saturation. The use of Schottky diodes (Sec. 1.19) reduces the transistor turnoff time to negligible proportions.

In Sec. 1.19 we considered how a diode bridged between collector and base of a transistor can serve to restrain a transistor from entering saturation (see Fig. 1.19-1). To review the matter, we note again that a transistor in saturation operates with a collector junction forward-biased to the extent of about 0.55 V or more. The transistor may therefore be kept out of saturation by the addition of a *diode clamp*, which does not allow the collector junction to become adequately forward-biased.

A transistor with a Schottky-diode clamp is indicated in Fig. 6.13-1a. Such a diode-transistor combination is referred to as a *Schottky transistor* and generally represented by the figure shown in Fig. 6.13-1b. The geometry (simplified and idealized) of an integrated-circuit Schottky transistor is shown in Fig. 6.13-1c. Observe that the metal contact to the *n*-type collector has been allowed to overlap the *p*-type base, thereby creating a metal-semiconductor (Schottky) diode junction between the metal and the collector. As discussed in Sec. 1.19, the base doping is graded so that the junction of the base and the metal is not rectifying.

There is available a family of high-speed TTL gates of which the schematic is as given in Fig. 6.12-1 but in which the input diodes are Schottky diodes and all transistors, except T_2 , are Schottky transistors. An exception is made of T_2 since, as we have noted, T_2 does not saturate. This Schottky family of gates has propagation times as low as 2 ns and rise and fall times in the range 2 to 3 ns. The power-supply current spikes are also reduced, being about 20 percent as large as those encountered in TTL gates that do not use Schottky transistors. In these gates the improved speed results, in part, from the fact that Schottky transistors have areas which are roughly one-half the areas used in conventional TTL transistors.

EMITTER-COUPLED LOGIC

6.14 OTHER LOGIC WITH TTL GATES

In Sec. 5.6 in connection with DTL NAND gates we noted the feasibility and advantages of the WIRED-AND connection. Such a connection, which directly ties the outputs of two or more gates together must be disallowed with TTL NAND gates. For consider that with the gates so joined, one of the gates is OFF while the other gate is ON. Then there would be a path from the supply voltage through a small collector resistor R_{c2} and through transistor $T2$ of the OFF gate into $T3$ of the ON gate. The current drawn would be approximately 40 mA, and instead of lasting a short time the current would last until the logic levels were changed. This would increase the drain on the power supply, but—most important—the dissipation in transistors $T2$ and $T3$ and the 100- Ω resistor would be excessive.

Still, the economy possible with the WIRED-AND connection is such that manufacturers find it advantageous to make available a modified TTL gate which allows such a connection. The modification consists in deleting the upper totem-pole transistor; i.e., these gates are intended to be used with an external passive pull-up (resistor), and, of course, the advantages of an active pull-up are thereby relinquished.

While it is possible to perform all logic functions, it is often very convenient and economical to have available an AND gate. Of course, two NAND gates may be cascaded to achieve the AND function. However, gates are available in which a second inversion is incorporated within the gate itself and at low power level.

NOR gates and gates which perform the AND-OR-INVERT (AOI) function are also available in TTL logic. The circuitry of these gates is somewhat different from the circuitry of NAND gates (see Probs. 6.14-1 to 6.14-4).

REFERENCE

- I Morris, R. L., and J. R. Miller, eds.: Designing with TTL Integrated Circuits, McGraw-Hill, New York, 1971.

INTRODUCTION

All the other logic forms considered (RTL, DTL, TTL) suffer from a common and fundamental limitation on their speed of operation. This limitation occurs because in all these logic types transistors are driven into saturation, resulting in an increased propagation delay time. This consideration prompts us to inquire whether a logic form is possible in which transistors are switched from cutoff to an operating point in the active region. Certainly the forms so far considered cannot be operated in this way; for the range of base-emitter voltages in the active region extends over only some tens of millivolts, and there are so many variable factors (temperature variations, variability due to manufacture, etc.) to contend with. Suppose we arranged that at one logic level a transistor stage is operating in its active region. Then a small drift of the applied input voltage would be enough to drive the transistor either to cutoff or saturation with a correspondingly large change in the stage output voltage.

It is possible to establish a transistor in its active region, with stability, by introducing negative feedback through the simple expedient of using a large emitter resistor. This is precisely what is accomplished when using an emitter

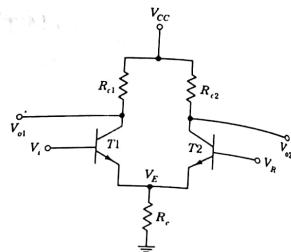


FIGURE 7.1-1
The difference amplifier.

follower or a phase splitter. The difficulty with this arrangement is that with a large emitter resistor, a large input-voltage swing at the base is required to carry the transistor from cutoff well into the saturation region. We can achieve both ends, however, i.e., active-region operation with stability and switching between this region and cutoff with a small input-voltage swing. These ends are accomplished by devising a circuit which will not turn a transistor current ON and OFF but will rather switch a current from one transistor to another.

The basic circuit configuration employed in the logic type under consideration, shown in Fig. 7.1-1, will, of course, be recognized as the *difference amplifier*, which was discussed in Sec. 2.1 in connection with its use in operational amplifiers and comparators. Since in this difference amplifier the emitters of the two transistors are connected, the logic family based on this circuit is referred to as *emitter-coupled logic*.

In the present application as a logic gate, the base of transistor T2 is held at a fixed reference voltage V_R while an input voltage V_i is applied to the base of transistor T1. When V_i is sufficiently lower than V_R , transistor T1 will be cut off and current will flow through T2. The reference voltage V_R and the resistors $R_{\text{E}1}$ and R_r are selected to assure that T2 operates in its active region and is not saturated. When V_i rises to equal V_R , the currents in the two transistors will be nominally equal. Finally, as V_i continues to increase, the emitter voltage V_E increases, since $V_E = V_i - V_{BE1}$ and V_{BE1} is approximately constant, and eventually T2 will cut off. We now have T1 operating in its active region. In summary, then, it appears that a variation of V_i will switch the current from one transistor to the other. As a matter of fact, we shall show that as V_i changes from the point where T1 is cut off to the point where T2 is just cut off, the total emitter current through R_r changes less than 2 percent. Thus, the mechanism of operation consists in switching a nominally fixed emitter current from one transistor to the other.

7.2 THE ECL GATE

An ECL gate, incorporating the basic structure of Fig. 7.1-1, is shown in Fig. 7.2-1. The input transistor T1 in Fig. 7.2-1 is shown here paralleled by a number of transistors to provide for multiple gate inputs.

Outputs are taken at the collectors through emitter followers. The emitter followers provide buffering and low impedance at the output terminals. An accurate, temperature-controlled reference voltage $V_R = -1.175$ V (at room temperature) is employed at the base of T2. The supply voltage employed is $-V_{EE} = -5.2$ V. Note that, contrary to the usual pattern, the collector resistors of the transistors are grounded and the emitter transistors are connected to a negative supply voltage. Thus, all the voltages encountered here will be negative. The reason for this reversal is discussed in Sec. 7.11.

As with other gates, ECL gates are commercially available in a number of types differing largely in the values of resistor components. Higher-resistance units dissipate less power and operate at a lower speed. Lower-resistance units dissipate more power but are faster. The gate with components and reference voltage in Fig. 7.2-1 is a Motorola unit of medium speed and power dissipation, designated as MECL II.

The ECL gate can operate as either a NOR gate or as an OR gate. It can be readily verified that if V_{o1} is taken as the output, a NOR gate is realized, while if V_{o2} is used, an OR gate results. In many an application either one or the other output will be used, while in some applications it is a great convenience to have available both outputs, which are, of course, complements of one another.

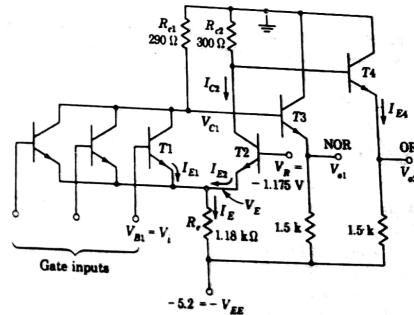


FIGURE 7.2-1
An ECL gate.

7.3 ECL-TRANSISTOR VOLTAGES

The saturation current of a transistor is $I_{\text{sat}} = [V_{\text{cc}} - V_{\text{c}}(\text{sat})]/R$, in which V_{cc} is the supply voltage and R is the total resistance in series with the transistor. If, as usual, $V_{\text{c}}(\text{sat}) \ll V_{\text{cc}}$, then $I_{\text{sat}} \approx V_{\text{cc}}/R$ rather independently of any transistor characteristic. The base-emitter voltage $V_{\text{be}} (= V_{\text{o}})$ required to cause this current I_{sat} to cross the emitter junction depends on the cross-sectional area of the emitter junction. That is, at least at the onset of saturation,

$$I_{\text{sat}} = I_{\text{c}} \approx I_{\text{e}} = I_{\text{eo}} e^{V_{\text{be}}/V_t} \quad (7.3-1)$$

Here I_{c} and I_{e} are the collector and emitter currents, respectively, and I_{eo} is the emitter-junction reverse saturation current, which is proportional to the area of the emitter junction. Hence, altogether V_{o} depends on V_{cc} , R , and the emitter-junction area.

For the gates considered so far it has turned out that the supply voltages, resistor values, and the geometry of the integrated transistor have been such that using $V_{\text{o}} \approx 0.75$ V gives reasonable agreement with measured voltages. For ECL gates, however, a better value is $V_{\text{o}} = 0.8$ V. This higher voltage value is a result principally of the smaller physical dimensions of the ECL transistor, a feature which serves as well to reduce capacitance and hence improve speed.

Consistent with previous procedures we may again consider that cutoff is at a voltage about 100 mV lower than V_{o} . On this basis we have $V_{\text{y}} = 0.70$ V. And, as has been our practice, we shall assume that when the transistor is in its active region, the base-emitter voltage, which we shall call V_{BEA} , is in $V_{\text{BEA}} = 0.75$ V, that is, midway between V_{y} and V_{o} . In summary, at room temperature, for the transistors in the ECL circuit of Fig. 7.2-1 we shall use

$$V_{\text{BE}} = \begin{cases} V_{\text{y}} = 0.70 \text{ V} & \text{cut-in} \\ V_{\text{BEA}} = 0.75 \text{ V} & \text{active region} \\ V_{\text{o}} = 0.80 \text{ V} & \text{saturation} \end{cases} \quad (7.3-2)$$

For other families of ECL gates corresponding voltages are slightly different. We shall neglect these differences. In any event, as usual, we take the temperature sensitivity of these voltages to be $-2 \text{ mV}/^{\circ}\text{C}$.

7.4 TRANSFER CHARACTERISTIC: THE OR OUTPUT

With all except one input transistor cut off, the transfer characteristic between V_{i} and V_{o2} (the OR output) is as given in Fig. 7.4-1a. We consider now how this characteristic comes about.

Referring to Fig. 7.2-1, we see that when V_{i} is sufficiently high, $T1$ will be ON and $T2$ will be OFF. Let us initially neglect the voltage drop through R_{c2} due to the base current of $T4$. Then $V_{\text{c2}} = V_{\text{b4}} = 0 \text{ V}$, and allowing a voltage of 0.75 V from base to emitter of $T4$, we find $V_{\text{o2}} = -0.75 \text{ V}$. However, as we shall now see, the neglect of the drop through R_{c2} is not entirely justifiable, and a small correction is in order.

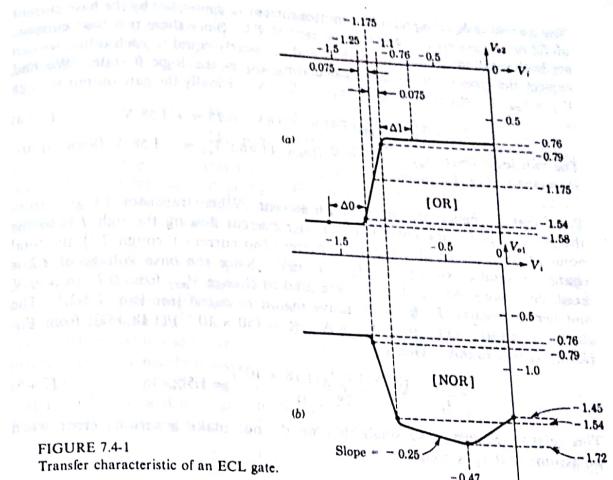


FIGURE 7.4-1
Transfer characteristic of an ECL gate.

The emitter current of $T4$ is

$$I_{\text{E4}} = \frac{-0.75 + 5.2}{1.5} \approx 3 \text{ mA} \quad (7.4-1)$$

Transistors in ECL gates have current gains h_{FE} in the range from about 40 to 150. We use $h_{\text{FE}} = 100$ as a typical value. The base current of $T4$ is $I_{\text{b4}} = I_{\text{E4}}/(h_{\text{FE}} + 1) = 3/101 \approx 0.03 \text{ mA}$. This base current flows through $R_{\text{c2}} = 300 \Omega$ and produces a voltage drop $300(-0.03 \times 10^{-3}) \approx 0.01 \text{ V}$. Hence $V_{\text{c2}} = V_{\text{b4}} \approx -0.01 \text{ V}$ rather than 0 V , and correspondingly $V_{\text{o2}} \approx -0.76$, as indicated in Fig. 7.4-1a.

Now let V_{i} decrease so that $T1$ turns off and $T2$ enters its active region. The emitter-to-ground voltage V_{E} is then

$$V_{\text{E}} = V_{\text{R}} - V_{\text{BEA}}(T2) = -1.175 - 0.75 = -1.925 \text{ V} \quad (7.4-2)$$

The emitter current is

$$I_{\text{E}} = \frac{V_{\text{E}} - (-V_{\text{EE}})}{R_{\text{e}}} = \frac{-1.925 + 5.2}{1.18} = 2.78 \text{ mA} \quad (7.4-3)$$

The current in R_{e2} is equal to the emitter current I_e , diminished by the base current of $T2$ and augmented by the base current of $T4$. Since these two base currents are both small in comparison with I_e and are nearly equal to each other, we can neglect the base currents when calculating V_{e2} in the logic 0 state. We find $V_{e2} = V_{ee} - R_{e2}I_e = -300(2.78) = -0.83$ V. Finally the gate output voltage is

$$V_{o2} = V_{ee} - V_{BE}(T4) = -0.83 - 0.75 = -1.58 \text{ V} \quad (7.4-6)$$

The two logic levels $V_{o2} = -0.76$ V (logic 1) and $V_{o2} = -1.58$ V (logic 0) are indicated in Fig. 7.4-1a.

Proof that I_e changes by less than 2.0 percent. When transistor $T1$ goes from the point of just being cut off (all emitter current flowing through $T2$) to the point where it carries all the emitter current (no current through $T2$), the total change in emitter voltage is $\Delta V_e = 50$ mV. Since the base voltage of $T2$ is fixed, this change ΔV_e is all that is required to change V_{BE2} from 0.75 to 0.70 V and hence to carry $T2$ from its active region to cutoff [see (Eq. 7.3-2)]. The change in emitter current is $\Delta I_e = \Delta V_e/R_e = (50 \times 10^{-3})/(1.18 \text{ k}\Omega)$; from Eq. (7.4-2), $I_e = 2.78$ mA. Hence

$$\frac{\Delta I_e}{I_e} = \frac{(50 \times 10^{-3})/(1.18 \times 10^3)}{2.78 \times 10^{-3}} = 1.5\% \quad (7.4-7)$$

This relative change is so small that we do not make a serious error when we assume that I_e is constant.

Transition width. As V_i increases, V_{o2} changes from -1.58 V (logic 0 assuming positive logic) to -0.76 V (logic 1). This change in logic level occurs as the input V_i swings through a transition region, which, as we shall now show, is 150 mV in width. We have

$$I_{E1} + I_{E2} = I_e = I = \text{const} \quad (7.4-8)$$

Over the transition region, $T1$ and $T2$ are both in their active regions; hence

$$I_{E1} \approx I_{E0} e^{V_{ee}/V_i} = I_{E0} e^{(V_{ee} - V_i)/V_i} \quad (7.4-9)$$

and $I_{E2} \approx I_{E0} e^{(V_i - V_{ee})/V_i} \quad (7.4-10)$

The ratio of emitter currents is

$$\frac{I_{E1}}{I_{E2}} = e^{(V_{ee} - V_i)/V_i} \quad (7.4-11)$$

Combining Eq. (7.4-9) with Eq. (7.4-6), we have

$$I_{E1} = \frac{I}{1 + e^{(V_{ee} - V_i)/V_i}} \quad (7.4-12)$$

and $I_{E2} = \frac{I}{1 + e^{(V_i - V_{ee})/V_i}} \quad (7.4-13)$

Suppose we define one edge of the transition region to correspond to the condition $I_{E1} = 0.05I$ and $I_{E2} = 0.95I$ while the other edge corresponds to $I_{E1} = 0.95I$ and $I_{E2} = 0.05I$. Then, as is easily verified, the total input voltage difference $\Delta V_{in} = \Delta V_i$, corresponding to the total width of the transition region,

$$\Delta V_i = V_{ee}(I_{E1} = 0.05I) - V_{ee}(I_{E1} = 0.95I) \\ \approx 2V_i \ln 20 = 6V_i = 150 \text{ mV} \quad (7.4-14)$$

As appears in Fig. 7.4-1, this transition range is symmetrically disposed with respect to the reference voltage V_R . The limits of the transition region occur at $V_i + 0.075$ and $V_i - 0.075$ V at -1.1 and -1.25 V. The corresponding values of V_{o2} are -0.79 and -1.54 V.

Noise margin. It is now of special interest to observe that the reference voltage and resistors in the ECL gate of Fig. 7.2-1 have been selected so that the gate output voltages symmetrically straddle the input-voltage transition region. We note that the mean of the output voltages is $\frac{1}{2}(-0.76 - 1.58) = -1.170$. This is very nearly equal to the reference voltage $V_R = -1.175$ V. As a result the noise margins are very nearly equal.

The output of a driving gate is -0.76 V at logic 1. We note in Fig. 7.4-1 that in order for a driven gate to recognize an input as being at logic 1, this input must not be less than -1.1 V. Hence the $\Delta 1$ noise margin is

$$\Delta 1 = -0.76 - (-1.1) = 0.34 \text{ V} \quad (7.4-15)$$

Similarly the $\Delta 0$ noise margin is

$$\Delta 0 = -1.25 - (-1.58) = 0.33 \text{ V} \quad (7.4-16)$$

It is to be noted that these noise margins are typical and not worst-case margins.

7.5 THE NOR OUTPUT

We have defined the edges of the transition region of the OR output as being the points where $I_{C2} \approx I_{E2} = 0.05I$ and $0.95I$, where I is the nominally constant current through R_c . Since the current $I = I_{C1} + I_{E2}$ is constant, $I_{C1} \approx I_{E1} = 0.05I$ when $I_{C2} \approx I_{E2} = 0.95I$ and vice versa. Hence, the transition points for the OR and NOR output occur for the same values of input V_i as shown in Fig. 7.4-1b. Further, if we neglect the small difference between R_{c1} and R_{c2} , the corresponding output voltages are also the same for OR and NOR outputs.

When V_i is low enough for $T1$ to be cut off, the output voltage is $V_{o1} = -V_{BE1}(T3) - I_{B1}R_{c1} = -0.75 - 0.01 = -0.76$ V, just as for the logic 1 level of the OR output. When, however, V_i has increased to the point where all the emitter current has transferred to $T1$, a further increase in V_i will result in a

further increase in I_{C1} and the output V_{o1} will continue to fall. With $T2$ cut off, the gain A from input to collector of $T1$ is (see Sec. 6.5)

$$A = \frac{\Delta V_{C1}}{\Delta V_i} = -\frac{R_{e1}}{R_e} = -\frac{0.290}{1.18} = -0.25 \quad (7.5-1)$$

Hence, as indicated in Fig. 7.4-1, V_{o1} falls with this negative slope, until transistor $T1$ begins to approach saturation.

We now estimate the input voltage V_i at which saturation begins to make itself felt. A transistor which is well into saturation has, as we have frequently noted, a collector-emitter voltage in the range 0.1 to 0.2 V. However, as shown in Fig. 1.10-1, when a transistor is just entering the region of saturation, the collector-emitter voltage is more nearly about 0.3 V. With 0.3 V between collector and emitter, the voltages V_{C1} and V_E are

$$V_{C1} = -(5.2 - 0.3) \frac{R_{e1}}{R_{e1} + R_e} = -\frac{4.9(0.290)}{1.47} = -0.97 \text{ V} \quad (7.5-2)$$

and

$$V_E = -0.97 - 0.3 = -1.27 \text{ V} \quad (7.5-3)$$

The output voltage V_{o1} and its corresponding input voltage are then

$$V_{o1} = V_{C1} - V_{BE3} = -0.97 - 0.75 = -1.72 \text{ V} \quad (7.5-4)$$

and

$$V_i = V_E + V_o = -1.27 + 0.8 = -0.47 \text{ V} \quad (7.5-5)$$

in which we have used 0.75 and 0.8 V, respectively, as the base-emitter drop for $T3$ (in the active region) and $T1$ (in saturation).

With a still further increase in V_i , the output V_{o1} begins to rise because additional emitter current in $T1$ is diverted to the base and away from the collector. When $V_i = 0$ V, $T1$ is deeply in saturation. Using 0.8 and 0.1 V, respectively, as the base-emitter voltage and the collector-emitter voltage of $T1$, we have $V_E = -0.8 \text{ V}$ and $V_{C1} = -0.8 + 0.1 = -0.7 \text{ V}$. The output voltage is then

$$V_{o1} = -0.7 - 0.75 = -1.45 \text{ V} \quad (7.5-6)$$

as indicated in Fig. 7.4-1b.

When in Fig. 7.2-1, $T1$ and some of its paralleling transistors conduct, the current in R_{e1} may be somewhat larger than the current in R_{e2} when $T2$ is ON. To reduce the corresponding difference in the logic 0 levels of OR and NOR outputs, R_{e1} is made smaller than R_{e2} .

7.6 MANUFACTURER'S SPECIFICATIONS: TRANSFER CHARACTERISTIC

The ECL circuit shown in Fig. 7.2-1 corresponds in all respects, component values, supply and reference voltages, etc., to the line of ECL logic manufactured by Motorola and designated MECL II. The plots of Fig. 7.6-1 are published by

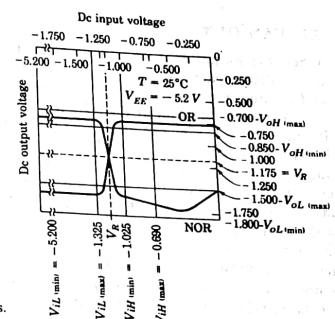


FIGURE 7.6-1
Typical ECL transfer characteristics.

Motorola as typical, average transfer characteristics for the gate. There is generally good agreement between the plots of Fig. 7.6-1 and the derived transfer characteristic given in Fig. 7.4-1.

The maximum range of variability to be anticipated in these gates is indicated in Fig. 7.6-1. Thus, while the available output voltage V_{oH} is nominally -0.76 V , the actual output may range from $V_{oH}(\max) = -0.70 \text{ V}$ to $V_{oH}(\min) = -0.85 \text{ V}$. Similarly the output voltage V_{oL} ranges from $V_{oL}(\max) = -1.5 \text{ V}$ to $V_{oL}(\min) = -1.8 \text{ V}$. The manufacturer specifies that an input voltage not more negative than $V_{iH}(\min) = -1.025 \text{ V}$ is guaranteed to be to the right of the transition region and hence to be acknowledged by the gate as corresponding to a logic 1 input. The ΔI margin, allowing for worst-case possibilities, is

$$\Delta I = V_{oH}(\min) - V_{iH}(\min) = -0.850 + 1.025 = 0.175 \text{ V} \quad (7.6-1)$$

rather than the typical value 0.34 V given in Eq. (7.4-10). Similarly, the worst-case ΔO margin is

$$\Delta O = V_{oL}(\max) - V_{iL}(\max) = 1.500 - 1.325 = 0.175 \text{ V} \quad (7.6-2)$$

The $V_{iL}(\min) = -5.2 \text{ V}$ is specified just to indicate that even if the most negative voltage available should happen to be impressed on the input, no disadvantage would result. The voltage $V_{iH}(\max) = -0.690 \text{ V}$ is so specified to indicate that a voltage in excess of this value will cause transistor $T1$ to carry an unnecessarily large current which will eventually produce saturation.

Notice again that it is required (to assure equal noise margins $\Delta O = \Delta I$) that the midpoints of the transition regions of the OR and NOR outputs occur at the coordinates $V_i = V_R$ and $V_{o2} = V_{o1} = V_R$.

7.7 FAN-OUT

When the output of a gate is at logic 0, it need furnish no input current to the driven gate. Input current is required when the logic level is logic 1, and the question of allowable fan-out arises.

We have seen that at logic 1 V_{o2} (or V_{o1}) is at -0.76 V when no current is being drawn. We have also noted that with $V_i = -0.76$ the ΔI noise margin is $\Delta I = 0.34$ V [see Eq. (7.4-12)]. If, then, we fanned out to a number of gates, the output voltage V_{o2} would fall below -0.76 V and there would be a corresponding reduction in the noise margin. Thus, in the present ECL case (as for the RTL gates as well) the allowable fan-out is a continuous function of what we decide is an acceptable noise margin. This situation is different from that which prevails in the DTL and TTL gates, where the outputs were taken across a saturated transistor. In those cases, as long as the allowable fan-out is not exceeded, the output voltage of the output transistor remains essentially constant and so does the noise margin.

A fan-out calculation is given in the following illustrative example.

EXAMPLE 7.7-1 The output V_{o2} of the gate of Fig. 7.2-1 is to be fanned out to N similar gates, as shown in Fig. 7.7-1. Find N at room temperature if the ΔI noise margin is to be 0.3 V. Assume the following worst-case conditions. The resistors of the driving stage are 20 percent higher than typical, $R_{o2} = 300(1.2) = 360 \Omega$; the emitter resistor $= 1.5(1.2) = 1.8 \text{ k}\Omega$; the resistors of the driven stages are 20 percent lower than typical, $R_e = 1.18(0.8) = 940 \Omega$. The supply voltage is 10 percent high, $V_{EE} = 5.2(1.1) = 5.7$ V. The transistors have current gains $h_{FE} = 40$. (These departures from typical values are all in the direction to reduce the fan-out.)

SOLUTION Refer to Fig. 7.7-1. At the edge of the transition region $V_i = V_{o2} = -1.1$ V, as indicated in Fig. 7.4-1. This voltage value is not affected by the change in R_o . If the noise margin is to be 0.3 V, we require that $V_i = V_{o2} = 0.3$, the value of R_e . Assuming, as usual, that $V_{BE1} = 0.75$, we have $V_E = -0.8 - 0.75 = -1.55$ V. $I_E = [-1.55 - (-5.7)]/940 = 4.4$ mA and

$$I_I = \frac{I_E}{h_{FE} + 1} = \frac{4.4}{41} = 107 \mu\text{A}$$

Turning now to the driving stage, we have $V_{B4} = V_{o2} + 0.75 = -0.8 + 0.75 = -0.05$ V. $I_{B4} = 0.05/360 = 139 \mu\text{A}$. $I_{E4} = (h_{FE} + 1)I_{B4} = 41(139) = 5.7$ mA. $I_4 = [-0.8 - (-5.7)]/1.8 = 2.7$ mA, so that

$$I_o = I_{E4} - I_4 = 5.7 - 2.7 = 3 \text{ mA}$$

The fan-out is

$$N = \frac{I_o}{I_I} = \frac{3.000}{107} = 28$$

The number $N = 28$ calculated in Example 7.7-1 is to be compared with $N = 25$ given by the manufacturers as a "dc fan-out." On the other hand, if

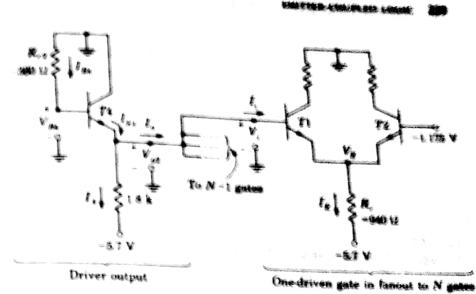


FIGURE 7.7-1
A calculation of fan-out.

we allow ΔI to fall to 0.1 V, then, as can be verified (Prob. 7.7-1), N becomes $N \approx 250$. In any event, it is apparent that fan-out in ECL gates is no problem if our only concern is the availability of enough driving current. However, the principal merit of ECL logic is its high speed. This speed is adversely affected by increasing the fan-out since each additional loading gate increases the loading capacitance correspondingly. Hence, the manufacturer specifies as well an "ac fan-out." This fan-out, usually about 15, is the fan-out recommended to keep rise and fall times and propagation times below specified limits.

7.8 SPEED OF OPERATION

Of all the gates considered, ECL is the fastest. When unloaded, a MECL II gate will exhibit a propagation delay time of the order of 4 ns. An even faster gate (MECL III) has a typical propagation delay time of about 1 to 2 ns. However, the speed of an ECL gate is adversely affected by capacitive loading. This situation is a result of the characteristic of the emitter followers used at the gate outputs. For, when the base voltage of an emitter follower changes sharply in the direction to increase emitter current, the emitter follows. Any capacitance hanging across the output charges rapidly through the low output impedance ($\approx 6\Omega$ in MECL II) of the emitter follower. When, however, the input changes in the reverse direction, the emitter voltage remains fixed momentarily due to the coupled capacitor. Since the base voltage has dropped, the emitter follower cuts off and the capacitance must discharge through the relatively large emitter resistance ($-1.5 \text{ k}\Omega$ in Fig. 7.2-1). However, as we can now see, in the present ECL case, if the capacitive loading is moderate, the discharge time need not be excessively large because the logic levels are separated by a voltage difference which is small in comparison with the separation between the logic levels and the supply voltage.

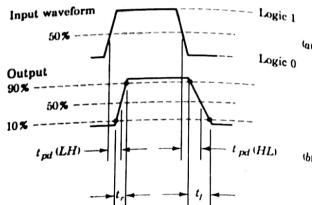


FIGURE 7.8-1
Propagation delay and rise and fall times.

For consider the on output of the ECL gate. Initially the output is at logic level 1 at $V_{o2}(1) = -0.76$ V. When the emitter follower ($T4$ in Fig. 7.2-1) cuts off, the output starts to fall, heading asymptotically toward $-V_{EE} = -5.2$ V. Logic level 0 is reached when the output voltage becomes $V_{o2}(0) = -1.58$ V. The time T required to make the transition between logic levels can be shown to be

$$T = RC \ln \left[\frac{V_{o2}(1) - (-V_{EE})}{V_{o2}(0) - (-V_{EE})} \right] \quad (7.8-1)$$

Using the values given above, we find that

$$T \approx 0.2RC \quad (7.8-2)$$

Assuming a capacitive load $C = 5$ pF, we find $T = 1.5$ ns.

Consider then a gate input as in Fig. 7.8-1a. With a substantial capacitive load on the gate output, the OR output would exhibit the characteristic to be noted in Fig. 7.8-1b, where the output falls more slowly than it rises. Propagation times in ECL gates are measured at a voltage midway between the two logic levels, the 50 percent point noted in the figures. We observe that the propagation time for a negative-going output $t_{pd}(HL)$ is longer than $t_{pd}(LH)$, the propagation time for a positive-going output.

In MECL II, it turns out that the input capacitance of a gate averages about 3.3 pF. Allowing for the capacitance associated with wired interconnections, we reasonably consider that each added fan-out contributes 5 pF. At no fan-out, we find that $t_{pd}(HL)$ and $t_{pd}(LH)$ are about equal, each being approximately 3.5 ns. However, at a fan-out of 20, corresponding to a capacitive load of $5(20) = 100$ pF, the manufacturer informs us that $t_{pd}(LH)$ has increased to only about 5 ns while $t_{pd}(HL)$ has increased to about 18 ns. Similarly, we find that at no fan-out, the rise and fall times t_r and t_f (see Fig. 7.8-1) are 6 and 4 ns, respectively. At a fan-out of 20, t_r increases to about 8 ns while t_f increases to about 30 ns. The times $t_{pd}(HL)$ and $t_{pd}(LH)$ can be reduced, of course, by shunting the emitter resistor of the output emitter

follower by an external resistor, at the expense, however, of increased power dissipation. In any event, it is apparent that if we wish to preserve the high speed inherent in ECL gates we must restrict the fan-out.

7.9 TEMPERATURE-COMPENSATED BIAS SUPPLY

In all of the preceding discussion we have assumed operation at a single temperature, 25°C. Of course, the transfer and other characteristics of ECL gates are temperature-dependent, as with other gates. And, as with the other gates, the principal source of the dependence is the temperature variation of the voltage drop across the forward-biased base-emitter junctions of the transistors.

There is one special point to be made in connection with the temperature variation of an ECL gate, which we shall now consider. We noted earlier that the ECL gate was designed so that the ΔI and $\Delta \theta$ noise margins should be approximately equal. If the reference voltage V_R is kept fixed, this condition can apply at only a single temperature. It is possible, however, to supply a reference voltage which is itself temperature-dependent so as to ensure that the symmetry of the noise margins is maintained over a wide range of temperatures. The bias supply circuit which accomplishes this for MECL II gates, and which is often incorporated directly on the integrated-circuit chip, is shown in Fig. 7.9-1. The operation of the circuit is analyzed below.

Assuming that diodes $D1$ and $D2$ operate at a forward bias of 0.75 V, at a temperature $T = 25^\circ\text{C}$, we find (neglecting the base current in $T5$) that the base voltage of $T5$ is -0.425 V. Assuming also a drop of 0.75 V from the base to emitter of $T5$, we find that $V_R = -0.425 - 0.75 = -1.175$ V, as expected.

Let us now assume a temperature change ΔT . In this case each forward-biased junction voltage changes by the amount $\delta = -k \Delta T$ ($k = 2 \text{ mV}^\circ\text{C}$). We calculate now the effect of such a change on the reference voltage V_R and on the output voltage levels $V_o(1)$ and $V_o(0)$ corresponding to logic 1 and logic 0

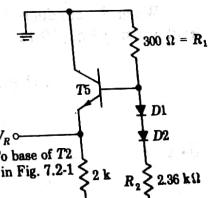


FIGURE 7.9-1
Reference supply circuit for ECL gates.

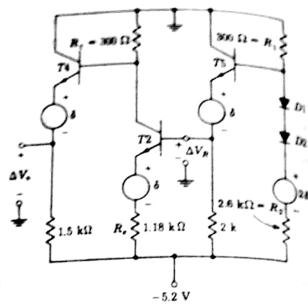


FIGURE 7.9-2
Equivalent circuit for calculating effect of temperature variation.

levels of the OR output (a calculation for the NOR output is left for the problems). The circuit of Fig. 7.9-2 includes the bias supply of Fig. 7.9-1 and that part of the gate circuit of Fig. 7.2-1 which generates the OR output. For ease of identification, the transistors in Fig. 7.9-2 have been given the same designations as in Figs. 7.2-1 and 7.9-1. Generators (δ and 2δ), representing the voltage increments introduced into the circuit due to a temperature change, have also been included.

Assuming that the gain through the emitter follower (T_5) is unity, we calculate the change in reference voltage V_R to be

$$\Delta V_R = \frac{2\delta R_1}{R_1 + R_2} - \delta = \delta \left(-1 + \frac{2}{1 + R_2/R_1} \right) \\ = \delta \left(-1 + \frac{2}{1 + 2.36/0.3} \right) \approx -0.77\delta \quad (7.9-1)$$

When T_2 conducts, the output is at logic level 0. The increment in this level is, using Eq. (7.9-1),

$$\Delta V_o(0) = -\Delta V_R \frac{R_c}{R_e} + \delta \frac{R_c}{R_e} - \delta \\ = (0.77\delta + \delta) \frac{R_c}{R_e} - \delta \\ = \left(1.77 \frac{0.3}{1.18} - 1 \right) \delta = -0.55\delta \quad (7.9-2)$$

Note, that in Eqs. (7.9-1) and (7.9-2) only resistor ratios appear. This feature is important because ratios of resistors can be held to much tighter tolerances

(≈ 2 percent) than the absolute values of resistors (≈ 20 percent). Finally when T_2 is cut off and the output is at logic level 1, the corresponding increment is

$$\Delta V_o(1) = -\delta \quad (7.9-3)$$

From Eqs. (7.9-2) and (7.9-3) we can now calculate that the average increment of the two logic levels is

$$\frac{\Delta V_o(1) + \Delta V_o(0)}{2} = \frac{-\delta - 0.55\delta}{2} \approx -0.77\delta \quad (7.9-4)$$

which is equal to the increment ΔV_R given by Eq. (7.9-1). Thus, as the temperature changes, the midpoint of the range from logic 0 to logic 1 changes by the same amount as the reference voltage V_R .

Since the reference voltage lies midway between the two output-voltage levels, independently of the temperature variation, the $\Delta 1$ and $\Delta 0$ noise immunities are the same. This therefore maximizes the noise immunity. However, the noise immunity does change with temperature (Prob. 7.9-2). Further, the above analysis neglected the effect of fan-out. It can be shown (see Prob. 7.9-3) that the result is relatively independent of fan-out.

The bias supply of Fig. 7.9-1 also provides a measure of compensation for variations in the supply voltage $-V_{EE}$. However, the compensation is not as exact as for temperature variations (see Prob. 7.9-5).

7.10 LOGIC VERSATILITY OF ECL GATES

As with DTL logic, it is possible to extend the logic capabilities of ECL gates simply by connecting the gate outputs together. For example, it can readily be verified that two or more emitter followers, operating into a common load,

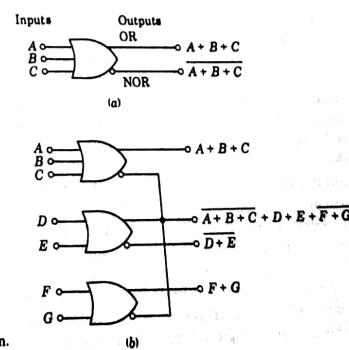


FIGURE 7.10-1
The WIRED-OR connection.

constitute an OR gate for positive logic. Hence, if the outputs of ECL gates are joined, this connection provides the logical sum (OR operation) of the outputs that would otherwise appear at the individual gates. Such a connection is therefore referred to as the WIRED-OR connection.

An example of the WIRED-OR connection is shown in Fig. 7.10-1. In Fig. 7.10-1a is shown the symbol for the ECL gate. Two outputs are shown. However, as a matter of convenience and to allow greater fan-out, some commercially available gates are provided with multiple OR and multiple NOR outputs. On the other hand, some gates are provided with only one OR and one NOR output. Figure 7.10-1b shows a WIRED-OR connection involving both OR and NOR outputs.

7.11 THE NEGATIVE SUPPLY VOLTAGE

In RTL, DTL, and TTL gates, the negative end of the power-supply voltage is grounded. In ECL gates, as we have noted, it is common practice to ground the positive end of the supply. We consider now the advantages of such an arrangement with ECL gates.

Let us initially put aside the question of grounding entirely and address ourselves to another matter. In Fig. 7.11-1 we have drawn an ECL gate. (To simplify the drawing we have omitted the part of the gate that provides the NOR output. This simplification will have no bearing on our discussion.) A supply voltage V_{CC} has been bridged across the gate. It has been our practice to consider that the output voltage of the gate is the voltage V_o taken between the emitter of $T4$ and the positive side of the power supply. But it readily appears that no fundamental change would be involved if we had chosen instead to take as the output voltage the voltage V'_o between emitter and negative side of the supply. As a matter of fact, as far as the signal, i.e., voltage changes, is concerned, the positive and negative sides of the supply are the same electrical point.

Next, we must recognize that closed-circuit loops (one of which is indicated in Fig. 7.11-1a) are formed by the connection to V_{CC} . Through these loops magnetic flux can thread, the flux being produced by the currents in the circuit shown or by currents in neighboring circuits. This flux, when changing, will produce an electromagnetic field in the loops in which V_{CC} is included. We should then really include in the power-supply loop (distributed) self-inductance and mutual inductance. Equivalently, we have chosen, in Fig. 7.11-1b to include instead, in series with the supply, a "noise" source V_n which is to represent all induced voltages. The characterization of V_n as noise is appropriate since these induced voltages will be random and unpredictable. A particular source of noise which V_n may represent is the following. Suppose that the power supply in Fig. 7.11-1 supplies not only the gate shown but also other circuits as well, which are paralleled across the supply. As these other circuits respond to the changing logic levels of their input signal, the current they draw from the

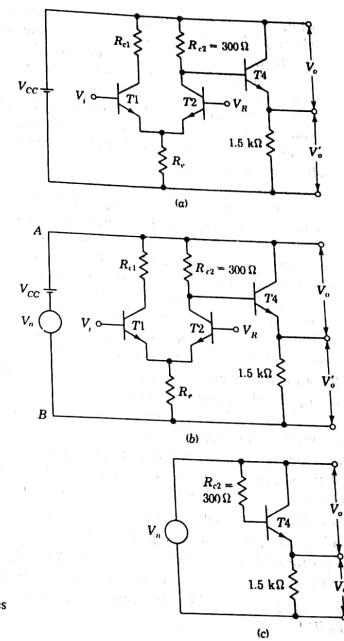


FIGURE 7.11-1
A negative supply configuration minimizes external-noise transfer.

power supply will change either transiently or more permanently. In any event, these current changes, flowing through the inductances or even through the finite impedance of the power supply itself, will generate voltages which may be represented by V_n .

It is now clear that the two sides of the power supply A and B are no longer equivalent and V_o and V'_o similarly are no longer equivalent. The voltages V_o and V'_o will reflect the noise to different extents. By way of example let us assume that $T2$ is cut off. Then the circuit to calculate V_o and V'_o is as appears in Fig. 7.11-1c. The impedance between collector and emitter of $T4$ is $R_{c2}/(h_{FE} + 1) = 300/100 = 3 \Omega$ for $h_{FE} = 99$. Hence, $V_o = (3/1,500)V_n = 0.002V_n$

while $V_s' = (1.500/1.503)V_s \approx V_s$. [A corresponding calculation for the case where T_2 is conducting is left as a problem (Prob. 7.11-2).] It is clear that the advantage lies with using V_s rather than V_s' .

It is well known that when a circuit is to operate in the proximity of a large mass of metal, it is advantageous electrically to connect this mass to some very common node in the circuit. The metal is then referred to as ground and provides a measure of shielding even though it need not physically intrude between the elements to be shielded from each other. This ground is often the chassis on the relay rack on which the circuit is built and will include the cabinet, if any, which houses the unit. It is also very common practice (though not necessarily universal) to arrange for the output terminals of the signal sources and the input terminals of signal-measuring devices (such as cathode-ray oscilloscopes, etc.) to use the ground as one signal terminal. This arrangement has the advantage that when units are interconnected, the chassis, cabinets, etc., can all be joined electrically.

Finally, returning to the ECL gate, it appears that we are initially at liberty to ground either the positive or negative side of the supply. Either would provide equivalent shielding. However, we find an advantage in using the positive side of the supply as one of the output terminals and a further advantage in using ground as one such terminal. Hence, altogether the practice is to ground the positive side of V_{cc} .

A second, less sophisticated reason for preferring the grounding arrangement normally employed with ECL is the possibility of an accidental short circuit developing between the output of a gate and ground. With the positive end of the supply grounded, as in Fig. 7.11-1, such a short places the entire 5.2-V supply across the 1.5-k Ω emitter resistor of the output emitter follower. The gate is able to tolerate such a short. On the other hand, with the negative end of the power supply grounded (Fig. 7.11-1), a short to ground from the gate output would place the entire supply voltage across the output transistor and at the same time apply the entire supply voltage to the transistor base through R_{e1} or R_{e2} (290 or 300 Ω). Such a short would promptly overheat and burn out transistor T_4 .

7.12 LEVEL TRANSLATION

It is often necessary to interconnect two different logic systems such as ECL and TTL (or DTL). One such application is the time-division multiplexing of M digital signals to form a single digital signal. Although the bit rate of each of the M signals may be handled using TTL, the bit rate of the composite signal is M times faster and may require ECL to process it.

Saturated logic to ECL translation The circuit of a commercial unit used to translate from saturated logic (TTL or DTL) to ECL is shown in Fig. 7.12-1. The circuit involving T_1 , T_2 , and T_4 is recognized as the ECL gate in which

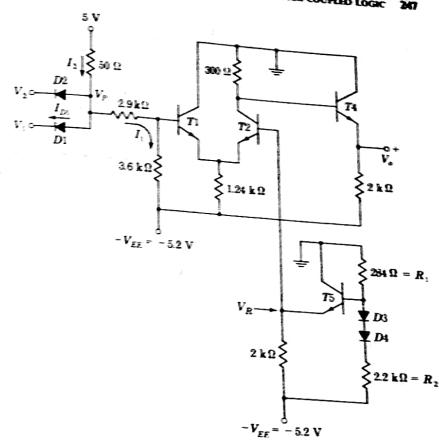


FIGURE 7.12-1
Saturated logic-to-ECL translator.

the transistor used to provide the NOR output (T_3 in Fig. 7.12-1) has been omitted. The component values are slightly different from those which appear in the circuit of Fig. 7.2-1. The collector resistor of T_1 has been omitted since no signal is taken from the collector. This omission has the advantage of assuring that T_1 will never saturate. The circuitry associated with T_5 provides the temperature-compensated reference voltage. When a single input variable is to be handled, one of the input diodes is returned to 5 V and is thereby cut off. If two logical variables are applied, then, as is readily verified, in addition to level translation the circuit performs the AND operation. The saturated logic levels of V_1 and V_2 are $V_{sl} \approx 0.2$ V and $V_{sh} \approx 3.5$ V (for TTL; DTL is generally higher). It is left as a student exercise (Prob. 7.12-1) to verify that these voltages, corresponding to logic 0 and logic 1, respectively, will produce, at the base of T_1 , voltages which correspond to logic 1 and logic 0 in ECL. Actually, the translation is accomplished by the input diodes and the three resistors bridged between +5 and -5.2 V. The rest of the circuitry provides temperature compensation and buffering to allow large fan-out.

The resulting transfer characteristic of the translator shown in Fig. 7.12-1 is given in Fig. 7.12-2.

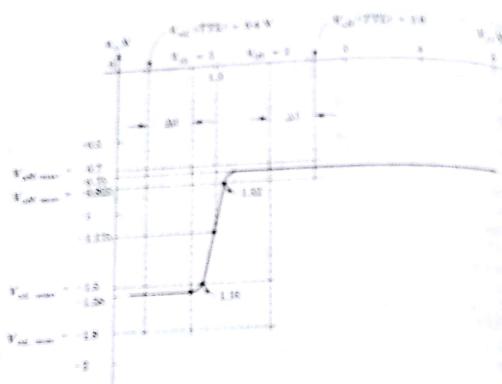


FIGURE 7.12-2
Transfer characteristics of translator shown in Fig. 7.12-1. $V_{EE} = 1\text{ V}$ and $V_{CC} = 2\text{ V}$
are manufacturer's specifications for the translator.

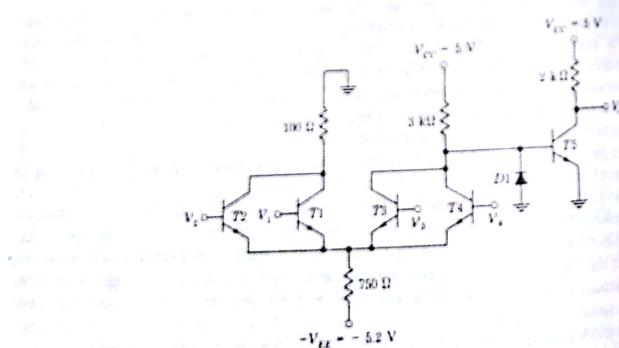


FIGURE 7.12-3
ECL-to-saturated-logic translator

ECL-to-saturated-logic translation. An ECL-to-saturated-logic translator is shown in Fig. 7.12-3. The circuit shown is capable of operating in either the on or saturated logic mode. To provide the logic operation V_2 is connected to the reference voltage V_3 , and V_3 is returned to logic 0. Then $V_2 = V_1$ since V_1 is returned to logic 0. $V_1 = V_2$ or V_1

The operation of this translator is rather straightforward; the translation being performed in the collector of transistors T^3 and T^4 since the collector resistor is returned to 5 V rather than ground. Diode D_1 is inserted, as in TTL, to damp out any ringing produced in the circuit. The details of these calculations and the determination of the transfer function are left for the problems.

7.13 ECL-GATE INTERCONNECTIONS

A pair of wires, a length of coaxial cable, etc., used to make interconnections between terminal pairs must, in principle at least, be viewed as a length of transmission line. The propagation of signals on transmission lines is considered in Appendix A. The transmission-line character of the interconnection makes itself especially apparent when the waveforms encountered make transitions between levels in times which are comparable to the time of propagation along the line. When the transition times are long in comparison to the time of propagation along the line, the line can be approximated by lumped-circuit elements.

Consider then, the connection of the output of a driving ECL gate to the input of a driven gate. The emitter-follower driver has a low output impedance ($< 10\ \Omega$) while the input impedance of the driven gate may well be of the order of many thousands of ohms. In Appendix A, we consider the characteristic impedance of interconnecting wires of geometries such as might reasonably be found in electronic circuitry. We estimate that such impedance would be in the range of some tens to some hundred of ohms. Hence, as indicated in Fig. 7.13-1a, the interconnection between gates may reasonably be represented by a line of one-way delay t_d and characteristic impedance R_0 . The sending-end termination is $R_s \ll R_0$, and the receiving-end termination is $R \gg R_0$.

If the input to the line V_i makes an abrupt transition between voltage levels, as indicated by the dashed waveform in Fig. 7.13-1b, the output V_o has the damped oscillatory waveforms shown by the solid plot. It is thus apparent that a transition at the driver-gate output from logic 0 to logic 1 might be interpreted at the driven gate as a sequence of several such transitions. If, on the other hand, the input waveform makes its transition in a time which is rather long in comparison with t_d , then, as shown in Fig. 7.13-1c, the output follows the input more closely; the oscillations of V_o about V_i are not nearly so pronounced.

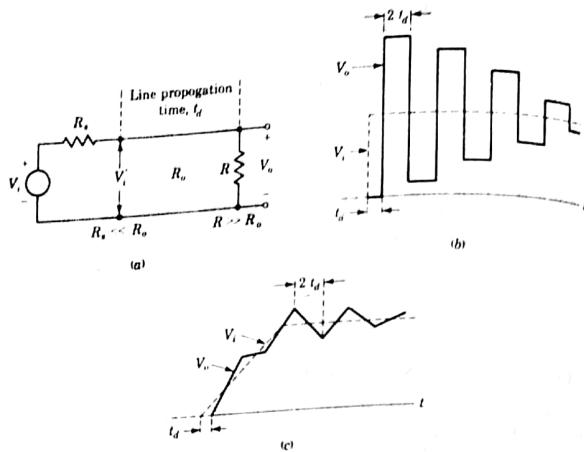


FIGURE 7.13-1
(a) A line of delay time t_d and characteristic impedance R_0 . It is terminated at its sending end in $R_s \ll R_0$ and at its receiving end in $R \gg R_0$. (b) The response at the output for a step input. (c) The response for an input ramp which rises in a time long in comparison with t_d .

We have already noted oscillations of the type indicated in Fig. 7.13-1b in connection with TTL gates (see Sec. 6.12). In ECL the problem is even more urgent because of its higher speed. Transition times between logic levels of order of 1 ns may occur in ECL. Depending on the dielectric constant of the insulation used to support the lines, 1-ns delays may occur in lines only 4 to 6 in long.

The oscillations can be suppressed by terminating the line at its receiving end in its characteristic impedance. In the absence of such termination, for an input logic swing of fixed rise time, the oscillation becomes more pronounced as the line gets longer. For this reason manufacturers generally provide guidance concerning the maximum allowable unterminated line lengths. This allowable length depends on the rise time, the fan-out, the characteristic impedance of the line, and the propagation time delay per unit length of line. By way of example, for Motorola type MECL 10,000, which has a rise time of 3.5 ns, using a microstrip line with $R = 50 \Omega$, and with a fan-out of 1, the allowable length is 8.3 in. For Motorola type MECL III, which has a rise time of 1 ns, using again a microstrip line except with $R_s = 100 \Omega$ and with a fan-out of 8, the allowable length is only 0.1 in.

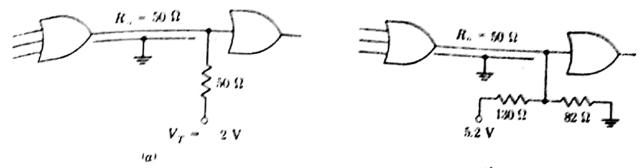


FIGURE 7.13-2
(a) A matched transmission-line interconnection between gates using an auxiliary supply voltage V_T for the termination. (b) An arrangement which avoids the

Here a 50- Ω line and matching termination have been used. As far as dc operation is concerned, one end of this terminating resistor is connected to the emitter of the output emitter follower of the driving gate. If this terminating resistor were larger, it would be allowable to return the other end to the -5.2-V supply. In such a case the terminating resistor would parallel the emitter resistor of the output emitter follower of the driving gate. In the present case, however, because of the small resistance of the termination, such a connection would result in excessive current through the output emitter follower, with consequent excessive dissipation in both transistor and termination. To circumvent this difficulty the terminating resistor is returned instead to an auxiliary terminating voltage V_T , which is commonly -2.0 V, as indicated in Fig. 7.13-2a. When the use to be made of the -2.0-V return is not adequate to justify a separate supply, the arrangement in Fig. 7.13-2b may be used. As can be verified, the -5.2-V supply and the 130- and 82- Ω resistors have a Thevenin equivalent replacement consisting of -2.0 V in series with 50 Ω . This latter arrangement limits the current in the output emitter follower but at the expense of considerable dissipation in the added resistors. Finally, we may note that in fast ECL gates, where it is virtually certain that a terminating arrangement as in Fig. 7.13-2 will be used, the manufacturer will often omit using an emitter resistor in the driving gate. In any event the output current diverted into the termination, and hence not available to drive gates, must be taken into account in estimating the allowable fan-out.

In Fig. 7.13-3 are shown a number of ways in which we can arrange the geometry of fan-out. In Fig. 7.13-3a all the driven gates are physically mounted very close to one another and are paralleled at the receiving end of a matched transmission line. In Fig. 7.13-3b the driving gate drives three transmission lines, two unmatched and one matched. The unmatched lines must be restricted in length in accordance with manufacturer's recommendations. In a representative case in an arrangement as in Fig. 7.13-3b we find with Motorola MECL III logic, using 50- Ω microstrip lines, the unterminated line with a fan-out of 1 may

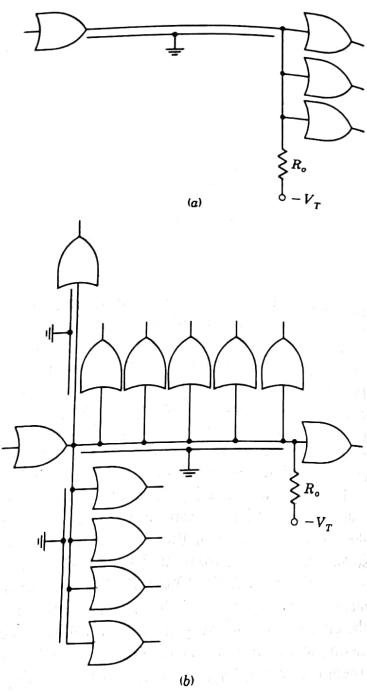


FIGURE 7.13-3
Fan-out arrangements for ECL gates: (a) the driven gates lumped at the end of a matched line; (b) gates distributed along matched and unmatched lines.

the load gates distributed nominally uniformly along the length of the line, the total length of the line may be no longer than 0.7 in.

In Fig. 7.13-3b the loads supplied by the matched line are distributed along the length of the line. A special consideration to be taken into account in connection with such an arrangement is discussed in the following illustrative example.

EXAMPLE 7.13-1 The matched line in Fig. 7.13-3b is 9 in long and drives 6 gates spaced at multiples of 1.5 in from the input end of the line. The line has capacitance per unit length and inductance per unit length $C = 2 \text{ pF/in}$ and $L = 0.02 \mu\text{H/in}$. The input capacitance of a gate is 5 pF. Estimate the value of the resistive termination required for the line.

SOLUTION Using Eq. (A.1-4), we find that the characteristic impedance of the line is

$$R_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.02 \times 10^{-6}}{2 \times 10^{-12}}} = 100 \Omega$$

The six gates have a total input capacitance of $6(5) = 30 \text{ pF}$. Since the length of the line is 9 in, the gates add a capacitance per unit length $C' = 30/9 = 3.3 \mu\text{F/in}$. If this added capacitance were uniformly distributed along the line, the line impedance would be given exactly by Eq. (A.1-4) with C replaced by $C + C'$. While such is not the case, we may well expect that this replacement will nonetheless yield a good approximation. We then find

$$R_0 = \sqrt{\frac{0.02 \times 10^{-6}}{5.3 \times 10^{-12}}} = 61 \Omega$$

The propagation delay per unit length [the inverse of the velocity in Eq. (A.1-5)] is

$$t_{pd} = \sqrt{L(C + C')} = \sqrt{(0.02 \times 10^{-6})(5.3 \times 10^{-12})} \approx 0.3 \text{ ns/in}$$

Series termination A line terminated at its receiving end is referred to as a *parallel-terminated line* since, as can be seen in Fig. 7.13-1, the matching resistor R is bridged across the line. A line may alternatively be matched at its input end by selecting $R_s = R_0$ in Fig. 7.13-1, in which case the line is referred to as a *series-terminated line*. It is feasible to provide such input matching, for the output impedance of a gate (see Fig. 7.2-1) is approximately $R_{c2}/h_{FE} = \frac{300}{30} = 6 \Omega$. We have noted that the line impedances are in the range 50 Ω to several hundred ohms. Hence, matching allows for the insertion of an additional resistor R_s in series with the output of the driving gate, as indicated in Fig. 7.13-4. Observe that series matching has the advantage that no auxiliary supply voltage is required. On the other hand, the series resistor limits the available output current and thereby restricts the fan-out to about 10.

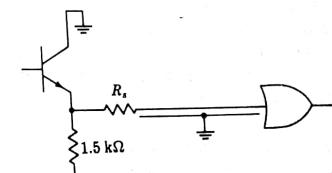


FIGURE 7.13-4
A line matched at its input end.

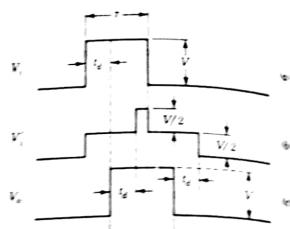


FIGURE 7.13-5
(a) A pulse of duration τ is applied to a line matched at its input end and open at its receiving end. Also shown is the waveform (b) at the input end and (c) at the receiving end of the line.

When driven gates are strung out on a parallel-terminated line, some difficulty may arise because signals from the driver arrive at different gates at different times. In series-terminated gates, of course, a similar situation prevails. However, an additional difficulty arises when employing the series-terminated line which does not occur in the parallel-terminated gate. Consider that in Fig. 7.13-1a, $R_s = R_o$, $R = \infty$, and V_i has the form indicated in Fig. 7.13-5a; that is, the driving gate changes its logic level, but the change persists only for a time τ . Suppose further that one of the driven gates is located at the input side of the

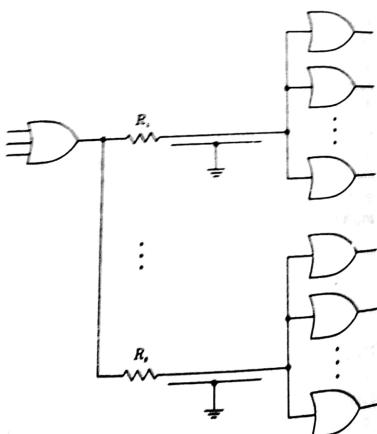


FIGURE 7.13-6
A number of series-matched lines used to accommodate a large fan-out

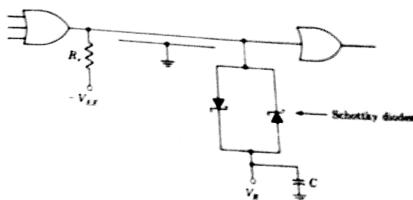


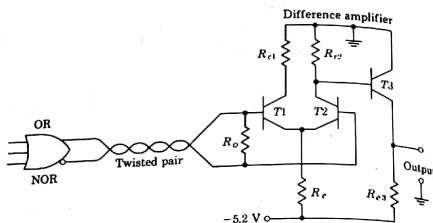
FIGURE 7.13-7
Diode termination of a line

line, where the voltage is V_i . It is left as a problem (Prob. 7.13-3) to verify that the input waveform and the waveform at the end of the line are given as in Fig. 7.13-5. We have drawn the waveforms for $\tau = 2.5t_d$, t_d being the one-way delay of the line. Note that the change in level in V_o persists for the same time τ as in V_i although there is a relative delay t_d . On the other hand, in V_i' , the full change persists only for a time $\tau - 2t_d$ and develops only after a time delay $2t_d$. If then it should happen that $\tau \leq 2t_d$, the full change in level would never appear in V_i' . The difficulty can be relieved by arranging the spacing between driven gates to be small in comparison with the spacing of the gates from the input side of the line. When many driven gates must be accommodated, the configuration of Fig. 7.13-6 is effective. Here a number of parallel lines are used, each with its own series termination.

Diode termination An additional method of suppressing oscillations is indicated in Fig. 7.13-7. This method is a rather natural extension of the use of input diodes, encountered earlier in connection with TTL gates. Of course, Schottky diodes are called for. We have noted that the ECL levels symmetrically straddle the reference voltage (see Fig. 7.6-1). Hence, the diodes are returned to this reference voltage V_R . Schottky diodes which have a cut-in voltage $V_c \approx 0.3 \text{ V}$ would allow the line voltage to swing freely through the range $V_R \pm V_c$, while oscillations outside this range would be sharply damped.

Twisted pair lines A second difficulty associated with the transmission of ECL waveforms is *crosstalk*, or unintended coupling of signals between circuits. Because of the speed of the signals, a large signal may be coupled from one signal path to another by a small stray capacitance or mutual inductance. Crosstalk can, of course, be minimized by using coaxial cables, but such cables are bulky and certainly do not readily allow a distribution of taps for driven gates.

A feature of ECL gates which is of great use in suppressing crosstalk is the fact that gates have two outputs (OR and NOR) which are of opposite polarity.

FIGURE 7.13-8
Twisted-pair transmission in ECL.

Whatever the change in voltage at one output, the change at the other output is equal and opposite. We take advantage of this feature as indicated in Fig. 7.13-8. The difference in output between OR and NOR output is transmitted over a twisted pair of wires to a difference amplifier. This difference voltage is nominally twice as large in amplitude as the signal available from the OR or the NOR output separately. The difference amplifier is normally made available by manufacturers for the present purpose and is generally referred to as a receiver. The twisting of the transmission wires keeps the wires together and also regularly reverses their relative positions. Hence, any signal path which might have induced in it a signal from one of the wires in the pair may well be expected to have an equal and opposite signal induced by the other wire. Hence, crosstalk from the twisted pair to other signal paths may be expected to be minimal. Similarly, crosstalk of other signals to the twisted pair will be introduced into the difference amplifier as a common-mode signal and hence will largely be restrained from appearing at the single-ended output. Of course, as appears, the twisted pair must be matched at its receiving end. A twisted pair may be used for transmission over many feet and may be used to distribute commonly used signals (such as clock waveforms) to many points. Generally, at each point when the signal is to be picked off the pair, a receiver will be required.

REFERENCE

- Blood, W. R., Jr.: "MECL System Design Handbook," Motorola Semiconductor Products, Inc., Phoenix, Ariz., October 1971.

MOS devices, as well as bipolar junction transistors (BJT), find application in logic gates. In this chapter we discuss the operation of PMOS (*p*-channel), NMOS (*n*-channel), and CMOS (complementary-symmetry) gates. CMOS is rapidly becoming the most favored because of its lower power dissipation, shorter propagation delay, and shorter rise and fall times.

8.1 ANALYTIC EQUATIONS FOR MOSFETS

Within a MOSFET, by definition, the charge carriers move away from the source and towards the drain. Therefore, in an *n*-channel device, where the carriers are negative, the conventional direction of current flow within the device is from drain to source. Thus the drain is positive with respect to the source, i.e., V_{DS} is positive as is also the current I_{DS} . Typical characteristics of *n*-channel MOSFETs are shown in Fig. 8.1-1. In Fig. 8.1-1a and b the MOSFET characteristics refer to an enhancement device. In such a device there is no channel between source and drain at $V_{GS} = 0$ V. No drain-to-source current I_{DS} flows until the gate-to-source voltage exceeds a threshold voltage V_T . This threshold

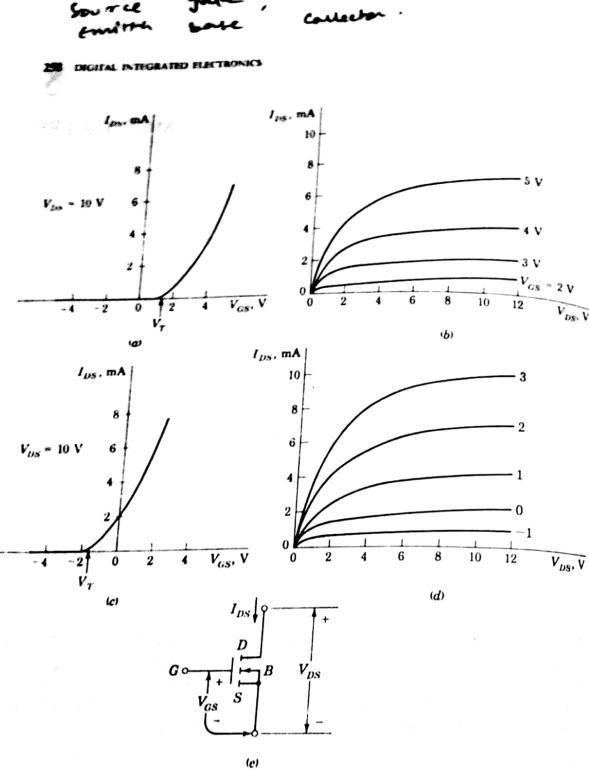


FIGURE 8.1-1
MOSFET characteristics. (a and b) Enhancement type. (c and d) Depletion type. (e) Defining voltages and currents. The device symbol represents an *n*-channel enhancement-type transistor.

voltage is of polarity which is the same as the polarity normally applied to the drain. Thus in an *n*-channel device, where V_{DS} is positive, so also is V_T , and a channel forms when $V_{GS} > V_T$.

In a *p*-channel device, where the carriers are positive, it is V_{SD} and I_{SD} (rather than V_{DS} and I_{DS}) which are positive. Furthermore, a channel forms to allow current I_{SD} to flow, when the source-to-gate voltage V_{SG} (rather than V_{GS}) exceeds a positive threshold voltage V_T , that is, when $V_{SG} > V_T$. We are using

HANDBOOK OF ELECTRONICS

the symbol V_T with two meanings. In an *n*-channel device V_T is defined as a particular value of V_{GS} , while in a *p*-channel device V_T represents a particular value of V_{SG} . Where any confusion may result we shall use instead the symbols $V_T(n)$ and $V_T(p)$. The symbolism we are employing avoids inconvenient negative signs and absolute-value signs.

In Fig. 8.1-1c and d typical characteristics are shown (again for an *n*-channel device) for a depletion transistor. Here a channel exists when $V_{GS} = 0$, and the threshold voltage V_T is negative. Strictly, such an *n*-channel transistor operates in the depletion mode when V_{GS} is negative and in the enhancement mode when V_{GS} is positive. It is customary nonetheless to refer to such a device simply as a depletion MOSFET. Both enhancement and depletion transistors are used in logic gates.

Either transistor type (enhancement or depletion) may operate in the non-saturation region (also referred to as the *triode* region in fond memory of the days of vacuum tubes) or in the saturation region. In the triode region there is a continuous channel between source and drain, and I_{DS} varies "linearly" with V_{DS} for fixed V_{GS} . At the source, the channel depth is nominally proportional to the extent to which the gate-to-source voltage V_{GS} exceeds the threshold voltage V_T and is thus proportional to $V_{GS} - V_T$. For fixed V_{GS} the channel depth is fixed. At the drain, the channel depth is proportional to the extent to which the gate-to-drain voltage V_{GD} exceeds V_T . Hence at the drain the channel off at the drain when $V_{GD} - V_T \leq 0$ when

$$V_{DS} \geq V_{GS} - V_T \quad (8.1-1)$$

When $V_{DS} \geq V_{GS} - V_T$, the transistor is in saturation. That is, because of the channel pinch-off, the current I_{DS} remains nearly constant, increasing only very slightly with increasing V_{DS} .

Just as we found it convenient to have analytic expressions for bipolar transistors (Ebers-Moll equations), so too is it useful to have analytic expressions for the MOSFET. In the triode region it is found that for an *n*-channel device

$$I_{DS} = k[2(V_{GS} - V_T)V_{DS} - V_{DS}^2] \quad 0 \leq V_{DS} \leq V_{GS} - V_T \quad (8.1-2)$$

In the saturation region

$$I_{DS} = k(V_{GS} - V_T)^2 \quad 0 \leq V_{GS} - V_T \leq V_{DS} \quad (8.1-3)$$

The constant k is given by

$$k = \frac{\mu e}{2t} \frac{W}{L} \quad (8.1-4)$$

where μ = mobility of carriers in channel (electrons in *n*-channel devices)

ϵ = dielectric constant of oxide insulating layer

t = thickness of oxide under gate

W = channel width

L = channel length

N MOS region

Typically, for *n*-channel devices $\mu_e/2t \approx 12 \mu\text{A/V}^2$ and for *p*-channel devices is smaller by about a factor of 3. The width-to-length ratio W/L may range from 0.1 for a load transistor to as high as 20 or 40 for a driver device. (See Sec. 1.13)

In a *p*-channel transistor, operating in the triode region the equations for the device current are more conveniently written in the form

$$I_{SD} = k[2(V_{SG} - V_T)V_{SD} - V_{SD}^2] \quad 0 \leq V_{SD} \leq V_{SG} - V_T \quad (8.1-5)$$

In the saturation region

$$I_{SD} = k(V_{SG} - V_T)^2 \quad 0 \leq V_{SG} - V_T \leq V_{SD} \quad (8.1-6)$$

These equations, like Eqs. (8.1-2) and (8.1-3), are approximations and do not include all effects which have an influence on device current; however, they are entirely adequate for our purposes of exploring the operation of FET logic gates.

8.2 TEMPERATURE EFFECTS

Equations (8.1-2), (8.1-3), (8.1-5), and (8.1-6) for the current I_{DS} (and I_{SD}) are affected by the temperature because both V_T , the threshold voltage, and the parameter k are temperature-sensitive. The temperature dependence of V_T is given approximately by

$$\frac{dV_T}{dT} \approx -2.5 \text{ mV/}^\circ\text{C} \quad (8.2-1)$$

The temperature sensitivity of k results almost entirely from the temperature sensitivity of μ [see Eq. (8.1-4)], the carrier mobility. The mobility decreases approximately inversely with the absolute temperature and hence so also does k . When there is a temperature increase, I_{DS} (or I_{SD}) increases because of the lowering of the magnitude of V_T and decreases because of the decreased carrier mobility. In a typical case we find that the effect of μ may be fivefold greater than the effect of V_T . The overall result is that generally the overall effect of a temperature increase is a decrease of current. In this respect the MOSFET differs from the bipolar transistor, where an increase in temperature increases the current both because the current gain h_{FE} increases and because the junction voltages decrease.

8.3 THE MOS INVERTER

As discussed in Chap. 1, the basic MOS switching-circuit configuration is an inverter which consists of a MOSFET switch driver driving a load which is itself a MOSFET device rather than a passive resistor. The driver is invariably an enhancement device since it is a great convenience that the driver be OFF when the gate voltage is at or near ground. When the driver is turned ON, it

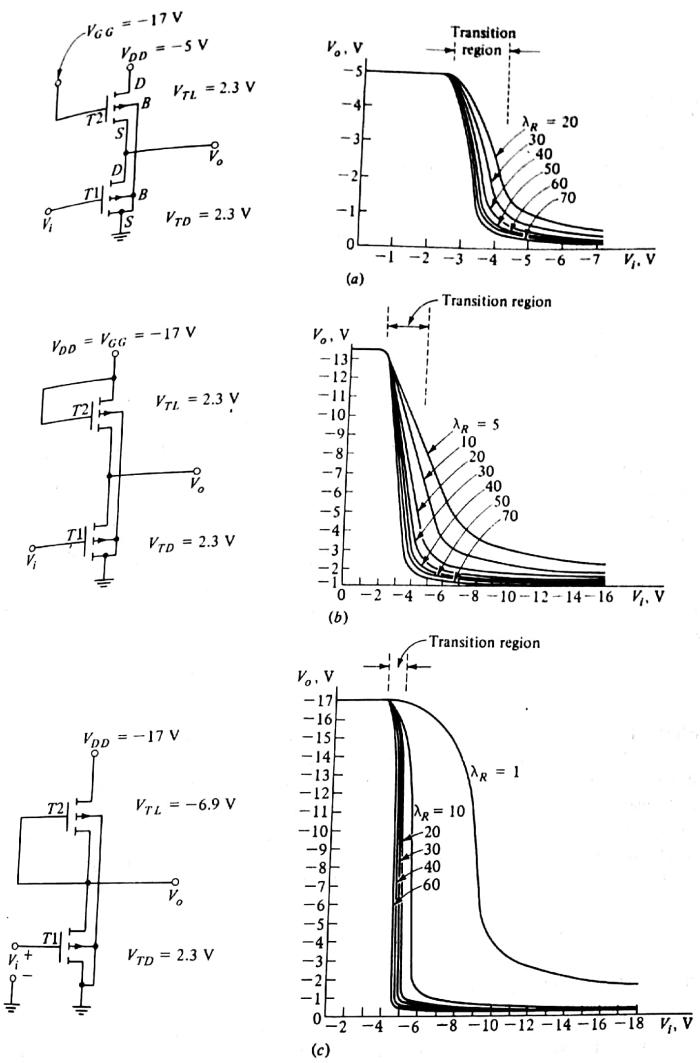


FIGURE 8.3-1
Input-output characteristics. (From "MOS/LSI Design and Applications," W. N. Carr, and J. P. Mize, McGraw-Hill, chap. 4, 1972).

invariably finds itself in the triode region. Such is the case since the gate voltage (furnished by another driving gate) will be at or near the supply voltage and the drain-to-source voltage will be at minimum magnitude. The load, on the other hand, may be an enhancement device or a depletion device and may operate in the triode or saturation region.

In Fig. 8.3-1 we display calculated input-output characteristics of PMOS inverters for three typical cases. In Fig. 8.3-1a both transistors are of the enhancement type, and both load and driver transistors have a threshold voltage $V_T = 2.3$ V. Since the transistors are *p*-channel devices, we apply the criterion given in Eq. (8.1-6) to determine whether we are in the triode or saturation region. Thus, to be in the triode region we require that $V_{SD} \leq V_{SG} - V_T$. This condition can be written as

$$V_{DG} \geq V_T \quad (8.3-1)$$

Since $V_{DG} = V_D - V_G = -5 - (-17) = 12$ V is greater than $V_T = 2.3$ V, the load transistor is biased to operate in the triode region. In Fig. 8.3-1b load the driven transistor remains as in Fig. 8.3-1a, but in this case the load transistor operates in the *saturation* region (the proof of this statement is left to the problems). In Fig. 8.3-1c the load transistor is a depletion device with a negative threshold voltage $V_T = -6.9$ V. The biasing of the load places it in the negative threshold voltage $V_T \leq -10.1$ V and in saturation when $V_o > -10.1$ V.

As we have discussed in Sec. 1.14 (see Fig. 1.14-1), we should expect the form of the input-output characteristic to depend principally on the parameter λ_R , defined by

$$\lambda_R \equiv \frac{\lambda_p}{\lambda_L} \equiv \frac{(W/L)_D}{(W/L)_L} \quad (8.3-2)$$

where $(W/L)_D$ = width-to-length ratio of channel in driver transistor

$(W/L)_L$ = width-to-length ratio for load

We noted that as λ_R increases, the transition of the output between its high and low levels becomes sharper. These expectations are confirmed in Fig. 8.3-1. In the calculations leading to the plots in Fig. 8.3-1 the effect of substrate bias is taken into account. Note also that the inverter using a depletion-mode MOSFET load has the steepest transition region.

8.4 THE CMOS INVERTER

The CMOS inverter is shown in Fig. 8.4-1a. The drains of a *p*-channel and an *n*-channel transistor are joined, and a supply voltage V_{SS} is applied from source to source. The output is taken at the common drain. The input V_i swings nominally through the range of V_{SS} . In the CMOS inverter shown, since we have grounded the source of the *n*-channel device, V_{SS} must be a positive voltage and V_i swings between ground and V_{SS} .

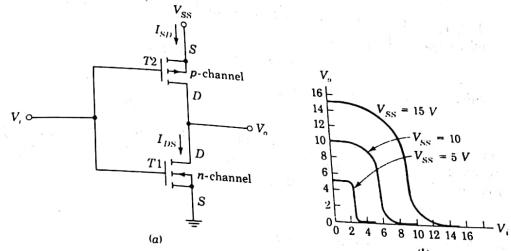


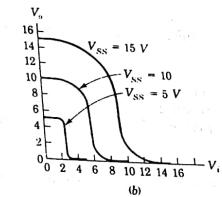
FIGURE 8.4-1
(a) A CMOS inverter and (b) its transfer characteristic.

Because of the complete symmetry of the circuit it seems intuitively clear that we shall want the two transistors to be reasonably alike. Therefore customarily it is arranged that the parameter k in Eqs. (8.1-2), (8.1-3), (8.1-5), and (8.1-6) are the same for the two transistors. The mobility of carriers in the *p*-channel device is smaller than the mobility in the *n*-channel device by a factor of 2 or 3. Hence to make the k 's equal, the ratio W/L for the *p*-channel must be correspondingly larger by a factor of 2 or 3 than W/L for the *n*-channel device [see Eq. (8.1-4)]. However, even with such an adjustment of the W/L ratio the CMOS inverter is not necessarily entirely symmetrical since the threshold voltages of the *p*-channel and *n*-channel devices generally turn out to be somewhat different.

Usually CMOS gates are designed to operate with supply voltages in the range 5 to 15 V. Typical transfer characteristics are shown in Fig. 8.4-1b. For the device to which Fig. 8.4-1b applies, the magnitude of the threshold voltage is about 2 V for each of the transistors. Observe the abruptness of the transition and that the total swing in voltage is equal to V_{SS} . In the MOS inverters such a situation prevails only when we arrange that the ratio λ_R of the λ 's be very large. In the present CMOS case, however, this situation prevails rather independently of the value of λ_R . Hence the CMOS inverter is often referred to as a *ratioless* inverter.

8.5 CALCULATION OF CMOS-INVERTER TRANSFER CHARACTERISTIC

It is instructive to use the device current equations to calculate the transfer characteristic of a CMOS inverter in a typical case. Referring to Fig. 8.4-1, we have that for $V_i \leq V_T(n)$, T1 is OFF, $T2$ is ON, and $V_o = V_{SS}$. Similarly for



$V_i \geq V_{ss} - V_T(p)$, T2 is off, T1 is on, and $V_o = 0$ V. Further, T1 is saturated when $V_{ds1} \geq V_{gs1} - V_T$, i.e.,

$$V_T(n) \leq V_i \leq V_i + V_T(n) \quad (8.5-1)$$

while T2 is saturated when $V_{sd2} \geq V_{gs2} - V_T$, i.e.,

$$V_o - V_T(p) \leq V_i \leq V_{ss} - V_T(p) \quad (8.5-2)$$

Thus, if, say, $V_T(n) = V_T(p) = 2$ V and $V_{ss} = 10$ V, we would have T1 saturated when $2 \leq V_i \leq V_o + 2$ and T2 saturated when $V_o - 2 \leq V_i \leq 8$. Or, to put the matter otherwise, T1 would be saturated when

$$V_o \geq V_i - 2 \quad (8.5-3)$$

and T2 would be saturated when

$$V_o \leq V_i + 2 \quad (8.5-4)$$

The currents I_{SD} and I_{DS} indicated in Fig. 8.4-1a are always equal. Accordingly, when T1 is in saturation and T2 is not, we have, using Eqs. (8.1-3) and (8.1-5),

$$k_n[V_i - V_T(n)]^2 = k_p[2(V_{ss} - V_i - V_T(p))(V_{ss} - V_o) - (V_{ss} - V_o)^2] \quad (8.5-5)$$

Here we have taken account of the fact that for the p-channel transistor $V_{SD} = V_{ss} - V_i$ and $V_{DS} = V_{ss} - V_o$. Similarly we find that when T2 is in saturation and T1 is not, we have

$$k_p[V_{ss} - V_i - V_T(p)]^2 = k_n[2(V_i - V_T(n))V_o - V_o^2] \quad (8.5-6)$$

Finally, when both transistors are in saturation, we find that

$$k_n[V_i - V_T(n)]^2 = k_p[V_{ss} - V_i - V_T(p)]^2 \quad (8.5-7)$$

Using Eqs. (8.5-5) to (8.5-7), we have plotted in Fig. 8.5-1 the input-output characteristic of a CMOS inverter for $V_T(n) = V_T(p) = 2$ V, $V_{ss} = 10$ V, and $k_p/k_n = 1$. Above the line $V_o = V_i - 2$ [Eq. (8.5-3)] T1 is in saturation. Below the line $V_o = V_i + 2$ [Eq. (8.5-4)] T2 is in saturation. In the region between the two lines both transistors are in saturation.

Note that the simultaneous saturation of both transistors defines a unique voltage $V_i(\text{sat})$, calculated from Eq. (8.5-7) to be

$$V_i(\text{sat}) = \frac{\sqrt{k_p/k_n}[V_{ss} - V_T(p)] + V_T(n)}{1 + \sqrt{k_p/k_n}} \quad (8.5-8)$$

In Fig. 8.5-1 with $k_p/k_n = 1$, $V_T(n) = V_T(p)$, $V_i(\text{sat}) = 5$ V. This voltage, at which there occurs an abrupt transition in output voltage is midway between 0 and V_{ss} because we have selected $k_p = k_n$. If k_p were not equal to k_n , then even if $V_T(n)$ were equal to $V_T(p)$, complete symmetry would not prevail. In any event we find from Eqs. (8.5-1) and (8.5-2) that the magnitude of the abrupt transition is given by

$$\Delta V_o = V_T(n) + V_T(p) \quad (8.5-9)$$

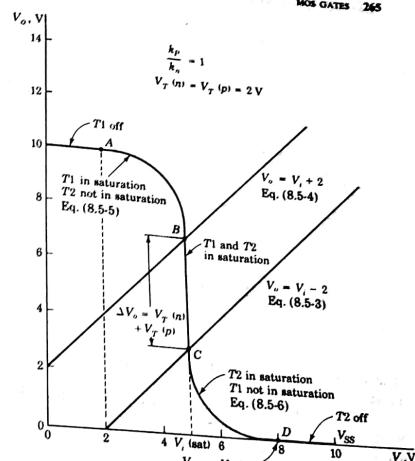
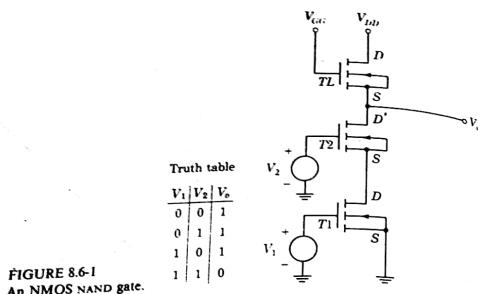


FIGURE 8.5-1
Transfer function of a CMOS inverter, $k_p/k_n = 1$.

The infinite slope displayed in Fig. 8.5-1 results from our assumption that in the saturation region the device current is absolutely independent of drain-to-source voltage, i.e., that the device is a constant current source. Such, of course, is not precisely so, and hence the transition from B to C in a physical situation would be sharp but not absolutely abrupt.

8.6 MOS GATES

Assuming that positive logic is intended, the NMOS circuit of Fig. 8.6-1 is a two-input NAND gate. The supply voltages V_{DD} and V_{GG} and the threshold voltage V_T are all positive. Logic 0 is represented by a voltage less than the threshold voltage and logic 1 by a voltage above the threshold voltage. The truth table given in Fig. 8.6-1 is readily verified. When either V_1 or V_2 or both are below threshold, only T_L conducts. In this case V_o is $V_{GG} - V_T$ or V_{DD} , whichever is lower. When both V_1 and V_2 are above threshold, both T1 and T2 conduct.

FIGURE 8.6-1
An NMOS NAND gate.

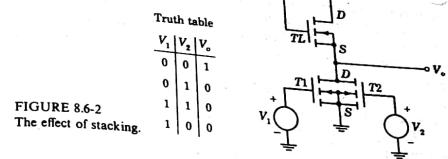
The output voltage V_o is the sum of the voltages across $T1$ and $T2$ and is presumably well below threshold.

Additional inputs can be provided by including additional devices in series with $T1$ and $T2$. The output V_o , when at logic 0, is the sum of the voltage drops across all the series devices. It is, of course, necessary to keep V_o at logic 0 and therefore comfortably below the threshold V_T . Hence as more series devices are included, the voltage drop across each one individually must be reduced. This is accomplished by increasing the width-to-length ratio W/L of the devices in order to reduce the resistance of the channel. Thus, suppose that starting with a design for a two-input gate, we wanted to modify the gate to accommodate three inputs. Then the two initial driver gates would be replaced by three driver gates, each with a W/L ratio three-halves the W/L ratio of the original devices.

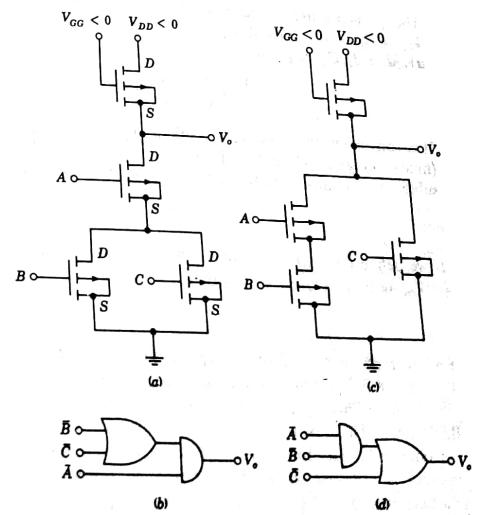
Since the same input voltage (with respect to ground) is on each gate input terminal, the gate-to-source voltage on each driven FET is not the same. The gate-to-source voltage is a maximum for the driver at the bottom of the stack and decreases as we go up the stack (see Fig. 8.6-1). As noted, driver transistors operate in the triode region where the device resistance is a function of gate voltage. Thus, we can compensate for this effect of stacking by making W/L progressively larger for transistors higher up the stack.

If the NMOS driver transistors are placed in parallel, as in Fig. 8.6-2, then, as can be verified, a NOR gate results.

If PMOS devices are used, the supply voltages V_{DD} and V_{GG} must be negative with respect to ground. It is left as a student exercise to verify (again for positive logic) that with drivers in series we have a NOR gate and with drivers in parallel we have a NAND gate. The feasibility of stacking MOS in series [which is not readily permitted with bipolar devices (see Prob. 8.6-3)] allows some

FIGURE 8.6-2
The effect of stacking.

interesting gate configurations like those shown in Fig. 8.6-3. Here we have used PMOS devices. The circuit in Fig. 8.6-3a accomplishes the logic indicated in Fig. 8.6-3b, and the circuit in Fig. 8.6-3c performs the logic indicated in Fig. 8.6-3d.

FIGURE 8.6-3
Illustrating the feasibility of stacking MOSFETs.

8.7 RISE TIME IN AN MOS GATE

A bipolar transistor is turned ON by establishing a distribution of minority-carrier charge in the transistor base and turned OFF by removing this minority charge. Additionally, capacitors must be charged and discharged. The speed with which bipolar transistor gates can be operated is therefore limited by the speed with which these operations can be performed. In MOS devices minority-carrier charge is not involved, and the speed of operation is determined only by the speed with which capacitors can be charged.

Consider then the situation represented in Fig. 8.7-1. This basic inverter becomes a multiple input gate if additional transistors are added in parallel or series with T_1 . The capacitance C_L represents the capacitance load on the gate and may well be the input capacitance of a succeeding gate. We inquire now into the rise time of the output voltage V_o ($= V_c$, the capacitor voltage) as T_1 is turned ON. We assume that V_c starts from 0 V. We consider first the case where the load transistor T_L operates in the saturation region, as would be the case if $V_{GS} = V_{DD}$.

In the saturation region we have, as in Eq. (8.1-3),

$$I_L = I_C = k_L(V_{GS} - V_T)^2 \quad (8.7-1)$$

The gate-to-source voltage of the load transistor with parameter $k = k_L$ is $V_{GS} = V_{DD} - V_c$. We therefore have that the rate of rise of V_c , which is $dV_c/dt = I_C/C_L$, is given by

$$\frac{dV_c}{dt} = \frac{I_C}{C_L} = \frac{k_L}{C_L}(V_{DD} - V_T - V_c)^2 \quad (8.7-2)$$

It is convenient to introduce the voltage $V_f \equiv V_{DD} - V_T$. This voltage V_f is the (final) voltage to which V_c will rise asymptotically. The solution of Eq. (8.7-2) subject to the initial condition that $V_c = 0$ at $t = 0$ is found to be (Prob. 8.7-1)

$$V_c = \frac{(k_L t / C_L) V_f^2}{1 + (k_L t / C_L) V_f} \quad (8.7-3)$$

If we define the rise time t_r to be the time required for V_c to rise from 0 V to $0.9V_f$, we find from Eq. (8.7-3) that

$$t_r = \frac{9C_L}{k_L V_f} \quad (8.7-4)$$

For example, if $C_L = 5 \text{ pF}$ (a typical input capacitance to a single gate) and $k_L = 20 \mu\text{A}/\text{V}^2$ (a typical value for a load transistor), and if $V_f = 10 \text{ V}$, we find $t_r \approx 0.2 \mu\text{s}$. This rise time is extremely long in comparison with the rise time in BJT gates. The reason for this slow response is basically the necessity for making the channel in the load transistor long and narrow. And this long narrow channel is required, as we have noted, in order to allow the input-output characteristic of the gate to exhibit an abrupt transition between logic levels. The long narrow channel limits the amount of current available to charge the capacitor C_L .

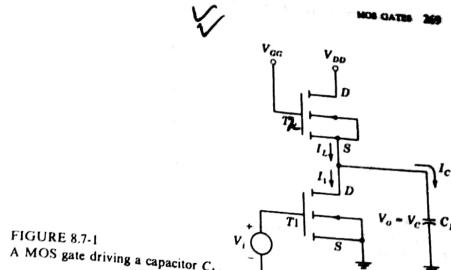


FIGURE 8.7-1
A MOS gate driving a capacitor C_L .

If, starting with a saturated load transistor, we increased the gate supply voltage V_{GS} to the point where $V_{GS} - V_T > V_{DD}$, the transistor would operate in the triode region. The increase in V_{GS} would increase the current through T_L and hence the current available to charge the load capacitor. However, a modest increase in V_{GS} would result in only a small increase in current, and no be limited basically by the restricted current available through the long narrow channel of the load. [The equation for $V_c(t)$ in the triode case is given in Prob. 8.7-2.]

8.8 THE FALL TIME

When, in the inverter of Fig. 8.7-1, the driver transistor T_1 is turned ON, the capacitor will discharge through T_1 and eventually fall nominally to ground. While T_1 is discharging, C_L , the current through T_L , continues in the direction to charge C_L . But, as we have seen, the W/L ratio of the driver is very much larger than the corresponding ratio for the load. Hence the discharge current through T_1 will be much larger than the charging current through T_L , and we shall therefore neglect the charging current.

Suppose then that T_1 , which is initially OFF, is turned ON by the application to its gate of a gating voltage $V_{GS} > V_T$. The volt-ampere characteristic of the transistor for this gating voltage is indicated in Fig. 8.8-1. Originally the transistor is OFF and operating at P_1 , where $V_{DS} = V_{CM}$, the maximum voltage drop across the capacitor. When T_1 is turned ON, the operating point moves abruptly to P_2 since the capacitor voltage cannot change instantaneously. The capacitor voltage $V_c = V_{DS}$ then decreases, and the transistor makes an excursion through the region of saturation as shown in Fig. 8.8-1. [We have idealized the saturation region to correspond exactly to Eq. (8.1-3), which assumes I_{DS} precisely constant and independent of V_{DS} .] At P_3 the triode region begins, and the

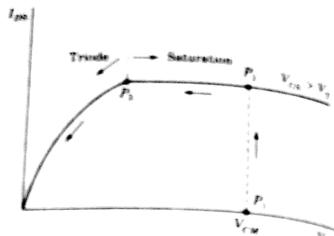


FIGURE 8.8-1
The operating path of T_1 as the capacitor C_L discharges

transistor and capacitor follow the triode characteristic, eventually to zero voltage.

Using Eq. (8.1-3), which applies in the saturation region, we find

$$V_C(t) = V_{CM} - \frac{I_{DS}}{C_L} t = V_{CM} - \frac{k_D}{C_L} (V_{GS} - V_T)^2 t \quad (8.8-1)$$

The point P_3 is reached when $V_{DS} = V_{GS} - V_T$. The time $t \equiv t_{sat}$ at which $V_C(t) = V_{GS} - V_T$ is calculated from Eq. (8.8-1) to be

$$t_{sat} = \frac{C_L}{k_D} \left[\frac{V_{CM} - V_{GS} + V_T}{(V_{GS} - V_T)^2} \right] \quad (8.8-2)$$

In the triode region, using Eq. (8.1-2), we find

$$\frac{dV_C}{dt} = -\frac{I_{DS}}{C_L} = -\frac{k_D}{C_L} [2(V_{GS} - V_T)V_C - V_C^2] \quad (8.8-3)$$

The time in the triode region required for $V_C(t)$ to fall from V_{CM} to $0.1V_{CM}$ is

$$t_{tride} = -\int_{V_{CM}}^{0.1V_{CM}} \frac{dV_C}{(k_D/C_L)[2(V_{GS} - V_T)V_C - V_C^2]} = \frac{1.15C_L}{k_D(V_{GS} - V_T)} \quad (8.8-4)$$

Let us take $k_D = 1 \text{ mA/V}^2$. Since we have assumed $k_L = 0.5 \mu\text{A/V}^2$, we have $k_D/k_L = 50$, which is quite reasonable for an MOS gate. Assume also that $V_{CM} = 10 \text{ V}$, $V_{GS} - V_T = 5 \text{ V}$, and $C_L = 5 \text{ pF}$. Then we calculate from Eqs. (8.8-2) and (8.8-4) that $t_{sat} = 0.004 \mu\text{s}$ and $t_{tride} = 0.016 \mu\text{s}$. The total fall time is then $t_f = 20 \mu\text{s}$. The fall time is thus very appreciably smaller than the rise time previously calculated to be $0.2 \mu\text{s}$. The principal reason for the large difference is the fact that $k_D \gg k_L$ because the channel of the driver is much wider and shorter than the channel of the load transistors.

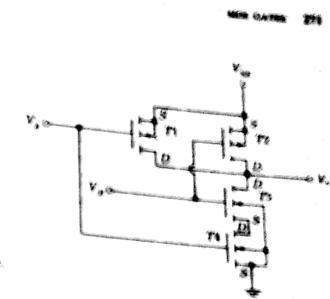


FIGURE 8.9-1
A CMOS NAND gate

8.9 THE CMOS GATE

A two-input CMOS NAND gate is shown in Fig. 8.9-1. Note that the driver transistors are series-connected while the load transistors are paralleled. The individual input is applied simultaneously to a pair of transistors, one driver and one load. Assuming positive logic and taking logic 0 to be nominally ground voltage and logic 1 to be nominally V_{DD} , we can easily verify that the circuit is indeed a NAND gate. The output V_O will be at logic 0 (ground) only when both inputs are at logic 1. The NMOS driver transistors are ON, in which case both PMOS load transistors will be OFF. The circuit in Fig. 8.9-2, in which the load devices are in series and the driver transistors in parallel, is readily verified to be a NOR gate.

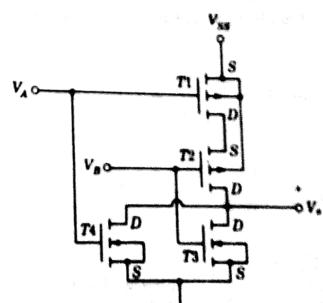


FIGURE 8.9-2

272 DIGITAL INTEGRATED ELECTRONICS

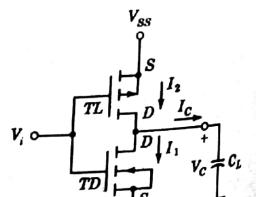
8.10 RISE AND FALL TIMES IN CMOS GATES

In NMOS and in PMOS the channel in the load devices must be very much longer and narrower than in the driver; i.e., we require $k_L \ll k_D$. This design is necessary to assure that when the driver is ON, the voltage drop across the driver transistor will be a very small fraction of the supply voltage. In CMOS, on the other hand, when the driver goes ON, the load is simultaneously driven OFF. Hence in CMOS, as already noted, $k_L \approx k_D$. Since the driver is *n*-channel and the load is *p*-channel, the *W/L* ratios are designed to be inversely proportional to the ratio of their mobility.

One of the major advantages of CMOS, then, is that there is always available a low-resistance channel path to charge and discharge a capacitive load across the gate output. Such a capacitive load at the output of a CMOS inverter is shown in Fig. 8.10-1. In an NMOS or in a PMOS gate the charging times of the load capacitor are widely different because of the difference in driver and load channel geometries, that is, k_D and k_L . In CMOS on the other hand, the load channel geometries are quite comparable.

The capacitor-charging calculations given above for the NMOS inverter can be applied directly in the present case. Referring to Fig. 8.10-1, suppose that *TL* has been ON and *TD* OFF, so that $V_C = V_{SS}$. The capacitor C_L will now discharge when the input gate voltage (from the output of a preceding gate) goes to $V_{GS} = V_{SS}$, thereby turning *TL* OFF and *TD* ON. The fall time t_f is to $V_{GS} = V_{SS}$, as given by Eqs. (8.8-2) and (8.8-4). In these equations $t_f = t_{sat} + t_{node}$, as given by Eqs. (8.8-2) and (8.8-4). In these equations $V_{GS}(\max) = V_{GS} = V_{SS}$. We may note that actually these equations apply somewhat exactly in the present case than in the case (of NMOS or PMOS) for which they were derived; for it will be recalled that in that derivation we neglected the small charging current through the load transistor while the capacitor was discharging. In the present case there is no current through the load device.

The parameter k for both the load and driver transistors will be comparable to the value of k ordinarily designed into the driver transistor of a gate without complementary symmetry. Hence in CMOS the fall time will be comparable to the fall time calculated in Sec. 8.7. However, because of the symmetry of the CMOS structure of Fig. 8.10-1 the rise time will be the same as the fall time if



$k_L = k_D$. As can be verified, except for the direction in which $V_E(t)$ is changing, the equations which described the charging of C_L are identical to the equations which describe its discharge.

8.11 MANUFACTURER'S SPECIFICATIONS

The specifications provided by manufacturers for CMOS devices are similar to those provided for BJT gates. These specifications deal with input and output currents and voltages, propagation delays, rise and fall times, etc. The specifications for the Motorola 4012 low-power NAND gate are given in Fig. 8.11-1. (Specifications for the type 4001 NOR gate are identical.) These specifications apply under the circumstances that a supply voltage $V_{SS} = 5$ V is employed. The definitions of the parameters V_{IH} , V_{IL} , V_{OH} , V_{OL} are given in Sec. 4.11. The current I_{IH} stands for the minimum current which must be supplied by a driving source if the CMOS-gate input is to be held at a voltage high enough, i.e., at V_{IH} or higher, for the gate to acknowledge that its input is at logic level 1. The other current symbols have similar meanings. Observe that the input currents, being only of the order of 10 pA, may generally be ignored. Referring to the output specifications, we note that the gate output is able to sink a maximum of 0.4 mA and still stay low enough in voltage to remain in the logic 0 region, that is, $V_o \leq V_{OL}$. At the other end, the gate is able to serve as a source of at most 0.5 mA and still remain in the logic 1 region.

Figure 8.11-1b shows the worst-case transfer characteristics of the gate. The transfer characteristic of a typical gate will lie somewhere between the worst-case limiting plots shown. With a 5-V supply the manufacturer specifies that

$$\Delta 0 = V_{IL} - V_{OL} \approx 1.5 \text{ V} \quad (8.11-1a)$$

$$\Delta 1 = V_{OH} - V_{IH} \approx 1.5 \text{ V} \quad (8.11-1b)$$

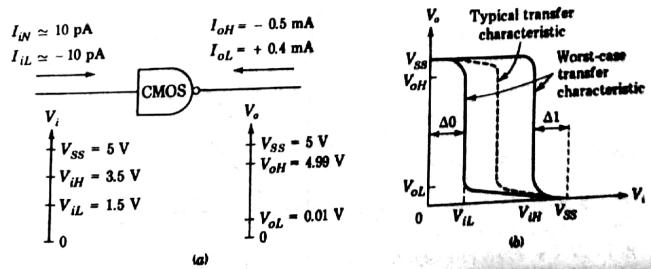


FIGURE 8.11-1 (a) Circuit connections and specifications. (b) Typical worst-case transfer characteristic.

Table 8.11-1

	Time, ns
$t_{pd} (LH)$	30
$t_{pd} (HL)$	30
t_r	60
t_f	60

If we inquire about typical rather than worst-case noise immunities, we find that with $V_{SS} = 5$ V, these are about 2.25 V. Thus the noise immunities are larger than those encountered in BJT gates with comparable supply voltages.

The propagation delay and transition times for the Motorola gates are given in Table 8.11-1, assuming a load consisting of a capacitor $C_L = 15$ pF in parallel with a resistor $R_L = 200$ k Ω . We assume that the input impedance of a CMOS gate is a 5-pF capacitance, so that this capacitive load represents a fan-out of 3. Increasing the fan-out increases the delay and transition times linearly, since the circuit time constant is directly proportional to the total load capacitance.

The CMOS gate referred to in Table 8.11-1 is relatively slow. However, high-speed CMOS gates are available having propagation delay times of the order of 20 ns. Recent advances in fabrication techniques, in which the substrate material employed is sapphire, have resulted in fast, low-power CMOS gates. These gates, called SOS/CMOS (silicon-on-sapphire) have rise and fall times

which are less than 20 ns. High-speed CMOS gates are today comparable to DTL gates, and a few claim to be as fast as the slower versions of TTL.

Typical quiescent power dissipation of these low-power CMOS units is 50 nW. However, at 100-KHz operation, the power dissipated is approximately 30 μ W. This dissipation increases at the rate of 20 dB/decade and is also a function of the capacitive load. Plots of the dissipation of the Motorola 4012 are shown in Fig. 8.11-2. (See also Sec. 1.15.)

Buffers The CMOS gate is a low-current gate designed to drive other CMOS gates. When a CMOS gate is to be used to drive a TTL or DTL gate, a CMOS buffer is often employed. An RCA 4009A buffer is capable of sinking a load current of 4 mA (I_{OL}) when it is in the low state and can source 1.75 mA when the output voltage is 2.5 V. This 2.5-V value is specified since it represents V_{IH} for many TTL gates.

The circuit configuration of the buffer is the same as the standard gate configuration. However, to obtain extra current capability, the dimensions of the CMOS devices used are increased.

8.12 INTERFACING BJT AND CMOS GATES

MOS gates and (at the present writing) CMOS gates are slower than BJT gates. On the other hand, the MOS gates can be fabricated with a component density on a silicon die which exceeds that possible with BJT devices. There is a merit to conserving "real estate" on the silicon die. For as the area involved on the die increases, so does the likelihood that a crystal imperfection will render the device defective, i.e., a reject. As a result there is an advantage in using both BJT and MOS devices in combination. The BJT devices are used where speed is required, and the MOS devices are used where slower operation is allowed. Since BJT and MOS devices generally operate at different voltage and current levels, some consideration must be given to proper interfacing. In many cases, interfacing requires interposing between the two types of devices circuits involving discrete bipolar transistors. Such would be the case, for example, if we needed to interface ECL logic operating between 0 and -5.2 V with MOS logic operating between 0 and +14 V. In simpler cases, interfacing may require a relatively minor accommodation in the BJT or MOS gate.

To illustrate some of the measures which may serve to effect a required accommodation we consider the interfacing between TTL logic and CMOS logic. This is the most common interfacing employed since TTL gates are the most often used logic types. TTL operates from a positive 5-V supply. We shall consider the case in which the CMOS as well operates from +5 V. For this purpose we have listed in Fig. 8.12-1 some of the relevant specifications for the TTL gate.

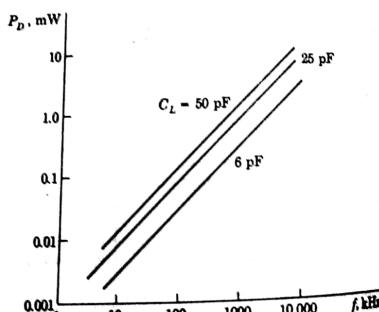


FIGURE 8.11-2
Power dissipation as a function of frequency.

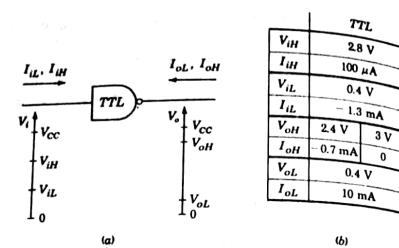


FIGURE 8.12-1
(a) Input and output characteristics. (b) Table of typical values.

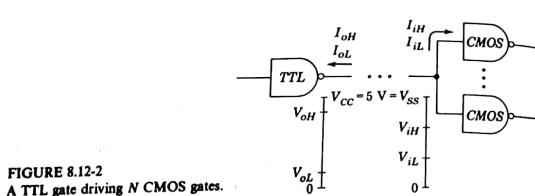


FIGURE 8.12-2
A TTL gate driving N CMOS gates.

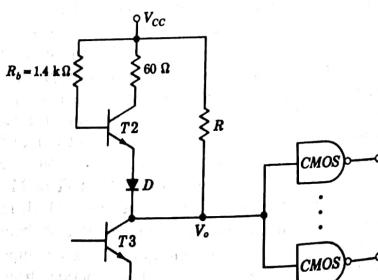


FIGURE 8.12-3
TTL with a passive pull-up.

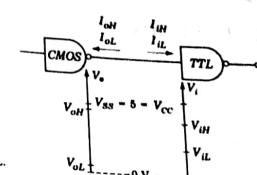


FIGURE 8.12-4
CMOS driving TTL.

We consider first the case of a TTL gate driving N CMOS gates, as in Fig. 8.12-2. Thus, the fan-out is N . For such an arrangement to operate successfully it is required that

$$-I_{oh}(\text{TTL}) \geq NI_{ih}(\text{CMOS}) \quad (8.12-1a)$$

$$I_{ol}(\text{TTL}) \geq -NI_{il}(\text{CMOS}) \quad (8.12-1b)$$

$$V_{oh}(\text{TTL}) \leq V_{il}(\text{CMOS}) \quad (8.12-1c)$$

$$V_{oh}(\text{TTL}) \geq V_{ih}(\text{CMOS}) \quad (8.12-1d)$$

As is readily verified from the data in Fig. 8.11-1 and in Fig. 8.12-1, Eqs. (8.12-1a) and (8.12-1b) are satisfied for any reasonable fan-out N . In addition, Eq. (8.12-1c) is also satisfied. However we find that Eq. (8.12-1d) is not satisfied since, even at "no load" $V_{oh}(\text{TTL}) = 3\text{ V}$ while $V_{ih}(\text{CMOS}) = 3.5\text{ V}$. A frequently employed circuit modification used to raise $V_{oh}(\text{TTL})$ above 3.5 V is shown in Fig. 8.12-3 where an external resistor R has been bridged between V_{cc} and the output. Typically R is in the range 2 to 6 k Ω .

When we consider a CMOS gate driving a TTL gate (see Fig. 8.12-4), we find that the condition $I_{ol}(\text{CMOS}) \geq -NI_{il}(\text{TTL})$ is not satisfied even for $N = 1$ since $I_{ol}(\text{CMOS}) = 0.4\text{ mA}$ while $-I_{il}(\text{TTL}) = 1.34\text{ mA}$. There are available, however, a number of CMOS buffers having adequate available output current (up to 6 mA)

REFERENCES

- Carr, W. N., and J. P. Mize: "MOS/LSI Design and Applications," McGraw-Hill, chap. 4, 1972.
- RCA Solid State Databook: SSD-203B, COS/MOS Digital Integrated Circuits.