CHAPTER 14.2
SWITCHES

Paul G. A. Jespers

THE IDEAL SWITCH

An ideal switch is a two-pole device that satisfies the following conditions:

Closed-switch condition. The voltage drop across the switch is zero whatever the current flowing through the switch may be.

Open-switch condition. The current through the switch is zero whatever the voltage across the switch may be.

Mechanical switches are usually electrically ideal, but they suffer from other drawbacks; e.g., their switching rate is low, and they exhibit jitter. Moreover, bouncing of the contacts may be experienced after closing, unless mercury-wetted contacts are used. Electronic switches do not exhibit these effects, but they are less ideal in their electrical characteristics.

BIPOLAR-TRANSISTOR SWITCHES

The bipolar transistor approximates an open switch between emitter and collector when its base terminal is open or when both junctions are reverse-biased or even only slightly forward-biased. Inversely, under saturated conditions, the transistor resembles a closed switch with a small voltage drop in series, typically 50 to 200 mV. This drop may be considered negligible in many applications.

Static Characteristics

A more rigorous approach to the transistor static characteristics is based on the Ebers and Moll transport equations

$$\begin{bmatrix} I_E \\ I_C \end{bmatrix} = I_S \begin{bmatrix} -\dfrac{1}{\beta_F} - 1 & 1 \\ 1 & -1 - \dfrac{1}{\beta_R} \end{bmatrix} \begin{bmatrix} \exp\,(V_E/V_T) - 1 \\ \exp\,(V_C/V_T) - 1 \end{bmatrix} \tag{1}$$

where V_E, V_C = voltage drops across emitter and collector junctions, respectively (positive voltages stand for forward bias, negative for reverse bias); $V_T = kT/q$ (typically 26 mV at room temperature); I_S = saturation current; and β_F, β_R represent forward (I_C/I_B) and reverse (I_E/I_B) current gains, respectively, with $V_E > 0$, $V_C < 0$ in the first case and $V_E < 0$, $V_C > 0$ in the second.

The saturation current I_S governs the leakage current flowing through the transistor under blocked conditions. It is always exceedingly small, and since it usually amounts to 10^{-14} or 10^{-15} A, it is difficult to measure. A standard procedure is to draw the plot representing the collector current in log scale versus the emitter forward bias V_E in linear scale. To find the saturation current, one must extrapolate the part of the curve, which can be assimilated to a straight line with a slope of 60 mV/decade to the intercept with the vertical axis for which $V_E = 0$. The saturation current can also be obtained with emitter and collector terminals permutated.

The current gains β_F and β_R can be evaluated by means of the same experimental setup. An additional ammeter is required to measure I_B.

It is common practice to rewrite Eq. (1) so that the emitter and collector currents are expressed as functions of

$$I_F = I_S[\exp(V_E/V_T) - 1] \tag{2}$$

and

$$I_R = I_S[\exp(V_C/V_T) - 1] \tag{3}$$

With these definitions Eq. (1) can be expressed as

$$I_C = I_F - I_R - I_R/\beta_R \tag{4}$$

$$I_E = -I_F - (I_F/\beta_F) + I_R \tag{5}$$

Hence, the Ebers and Moll transport model is found. This is illustrated by the equivalent circuit of Fig. 14.2.1. The leakage currents of the two diodes D_1 and D_2 are given respectively by I_S/β_F and I_S/β_R. With this model, it is possible to compute the currents flowing through the transistor under any circumstance.

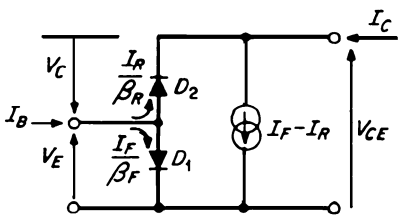

FIGURE 14.2.1 Equivalent circuit of the Ebers and Moll transport model of the bipolar transistor.

For instance, if the collector junction is reverse-biased and a small positive bias of, for example, +100 mV is established across the emitter-junction, the reverse current I_R is almost equal to $-I_S$ and I_F is equal to I_S exp (100/26) or 46.8 I_S. Hence, from Eqs. (4) and (5), the collector current is found to be equal to $48I_S$, and the emitter current is approximately the same with opposite sign. With the assumption that I_S is equal to 10 pA, both I_C and I_E are essentially negligible. *A fortiori*, I_B as derived from Eqs. (4) and (5) is also small:

$$I_B = (I_F/\beta_F) + (I_R/\beta_R) \tag{6}$$

To drive current through the transistor, the voltage across one of the two junctions must reach at least 0.5 V, according to:

$$V_E = V_T \ln(I_F/I_S) \tag{7}$$

or

$$V_C = V_T \ln(I_R/I_S) \tag{8}$$

derived from Eqs. (2) and (3).

The transistor operates in the saturation region when it approximates a closed switch. The voltage drop between the emitter and collector terminals is then given by

$$V_{CE,\text{sat}} = V_T \ln \frac{n + (n + \beta_F)/\beta_R}{n - 1} \tag{9}$$

where n represents the ratio $\beta_F I_B/I_C$, assumed larger than 1. For most transistors, this voltage drop lies between 50 and 200 mV. The inevitable resistance in series with the collector increases this voltage by a few tens of millivolts.

An interesting situation arises when I_C is almost equal to zero; e.g., when the bipolar transistor is used to set the potential across a capacitor. In this case, Eq. (9) becomes

$$V_{CE,\text{sat}} = V_T \ln (1 + 1/\beta_R) \tag{10}$$

Similarly, with interchanged emitter and collector terminals, the voltage drop is given by

$$V_{EC,\text{sat}} = V_T \ln (1 + 1/\beta_F) \tag{11}$$

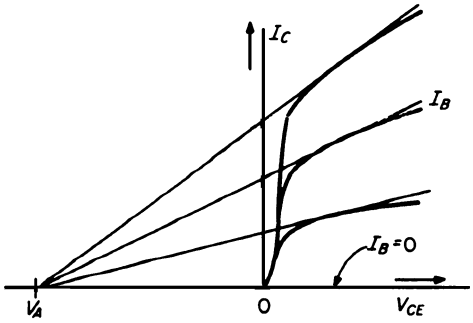

FIGURE 14.2.2 Typical common-emitter characteristics of the bipolar transistor. V_A is called the Early voltage.

Since β_F is normally much larger than β_R, $V_{EC,\text{sat}}$ may be smaller than 1 mV provided β_F is at least equal to 25. Consequently, inverted bipolar transistors are switches with a very small series voltage drop, provided that the current flowing through the transistor is kept small.

The two situations examined so far (open or closed switch) correspond in Fig. 14.2.2, respectively, to $I_B = 0$ and to the part of the curves closely parallel to the collector-current axis. The fact that all the curves have nearly the same vertical shape means that the series resistance of the saturated transistor is quite small. Since the characteristics do not coincide with the vertical coordinate axis, a small series emf must be considered, however, as previously stated.

A third region exists where the transistor plays the role of a current switch instead of a voltage switch. It concerns the switching from blocked conditions to any point within the active region or vice versa. Conceptually, the transistor may be compared to a controlled current source, which is switched on or off. However, because of the Early effect, the current is a function of the collector to emitter voltage. The Ebers and Moll model is inappropriate to describe this effect. A better expression of I_C is

$$I_C = I_S \exp (V_E/V_T)(1 + V_{CE}/V_A) \tag{12}$$

where V_A is called the *Early voltage.* Equation (12) is illustrated in Fig. 14.2.2. The finite output conductance of the transistor is given by I_C/V_A.

Dynamic Characteristics

The dynamic behavior of bipolar transistors suffers from a drawback called "charge storage in the base," which takes place every time transistors are driven in or out of saturation. The phenomenon is related to the majority carriers supplementing the minority carrier in the base to guarantee neutrality. The question is how to remove these extra carriers when the transistor is supposed to desaturate. Zeroing the base current is not a satisfactory solution for the majority carriers can only recombine with minority carriers. This requires a lot of time for lifetimes in the base are generally large in order to maximize the current gain. A better technique is to reverse the polarity of the base current. The larger this current, the more rapidly the majority carriers disappear and the faster the transistor gets out of saturation. Current continues to flow, however, until one of the two junctions gets reverse biased (usually the collector junction). Only then the transistor enters the active region and the collector current may start to decrease. When all majority carriers are swept away, both junctions are reverse-biased and the impedance of the base contact unfolds from a low impedance to an open circuit. A quantitative analysis of the desaturation mechanism can be found in the CD-ROM together with an example.

Charge storage is generally associated with the base of bipolar transistors, although the same phenomenon takes place near the depleted region in the emitter neutral region also. The reason why this is not considered is related to the doping profile in the emitter comparatively to the base region. The emitter doping is much larger

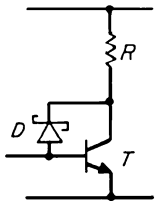

FIGURE 14.2.3 A Schottky diode *D* prevents *T* from going into saturation.

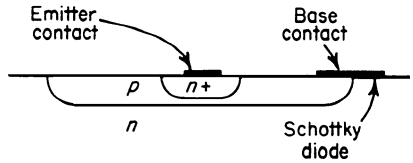

FIGURE 14.2.4 Planar *npn* transistor and Schottky diode in integrated-circuit form.

than the base doping because it increases the emitter's efficiency, which controls the current gain (the emitter efficiency is the ratio of emitter-base current over base-emitter current).

Junction diodes suffer from the same drawback. The diffusion length in the less-doped region being much longer than the base width of a bipolar transistor, charge storage problems should be worse. This is not the case, however, as diodes are generally bipolar transistors with their base and collector terminals shorted, making their charge storage effects similar to those of bipolar transistors.

In order to reduce the delay between the control signal applied to the base and the moment the collector current begins to decay, several techniques have been developed. Two of these are reviewed hereafter.

The first takes advantage of Schottky diodes, which exploit field emission and, therefore, ignore charge storage phenomena. Shottky diodes exhibit a slightly smaller voltage drop under forward bias than junction diodes (of the order of 0.4 V instead of 0.7 V), which is currently exploited to prevent bipolar transistors from getting saturated. The idea is illustrated in Fig. 14.2.3, which represents a bipolar transistor whose collector junction is paralleled by a Schottky diode. When the transistor is nearing saturation, the Schottky diode starts conducting before the collector junction does. This prevents the transistor from entering saturation. The base current in excess to what is needed to sustain the actual collector current flows directly to ground through the series combination of the forward-biased Schottky diode and emitter junction. Figure 14.2.4 shows how the combination of a Shottky diode and a bipolar transistor can be implemented. The Schottky diode consists of the metal contact that overlaps the lightly doped collector, whereas in the base region the metal to the P-type semiconductor resumes to an Ohmic contact. Such combination is currently used in Schottky logic, a family of fast bipolar logic circuits.

The second way to avoid charge storage is to devise circuits that never operate in saturation. The switch shown in Fig. 14.2.5, which involves two transistors controlled by a pair of square waves with opposite polarities is a good example. When Q1 is on, Q2 is off or vice versa. Current is switched either left or right. Although the circuit looks like a differential pair, it operates in a quite different manner: the conducting transistor is in the common base configuration, while the other is blocked. Since the output current is taken from the collector of the common base transistor, the output impedance is very large. This means that the switch is a current-mode

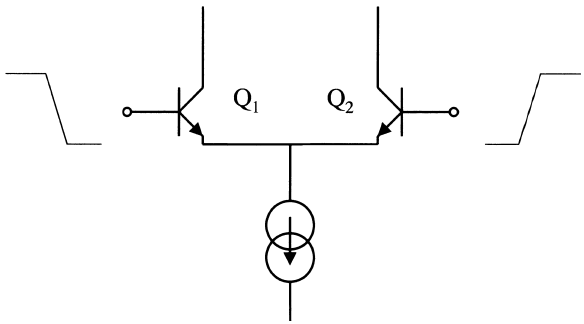

FIGURE 14.2.5 A bipolar current switch.

switch instead of a voltage-mode switch. Very short switching times are feasible this way for none of the two transistors ever saturates. Emitter-coupled logic (ECL) takes advantage of this circuit configuration.

MOS SWITCHES

Insulated gate field-effects (IGFETs, also called MOS transistors) and junction field-effect transistors (JFETs) can be put to use in order to mimic switches. They resemble an ideal closed switch in series with a linear resistor when "on" and an open switch when "off." The leakage current, however, is larger than with bipolar transistors.

Static Characteristics

When a field-effect transistor is turned on, its characteristics differ substantially from those of a bipolar switch. Since no residual emf in series with the switch is experienced, the transistor is comparable to a resistor whose conductance G is given by

$$G = \mu C_{ox} \frac{W}{L}(V_G - V_{T0} - \lambda V) \tag{13}$$

where μ is the mobility in the inversion layer, C_{ox} the gate oxide capacitance per unit area, V_G the gate-to-substrate voltage, V_{T0} the gate threshold voltage under zero bias, V the source or drain voltage, and λ a dimensionless factor that does take into account the so-called substrate effect (the value of λ lies somewhere between 1.3 and 1.5). The dependance on source and drain voltages of the conductance G represents a problem that impairs severely the performances of MOS switches. Consider, for instance, a MOS switch in series with a grounded capacitor to implement a simple sample-and-hold circuit. Since the MOS transistor is similar to a resistor when its gate voltage is high, the circuit may be assimilated to an RC network with a time constant that varies like the reciprocal of the difference in the right part of Eq. (13). Hence, when the input voltage V equals $(V_G - V_{T0})/\lambda$ the conductance becomes equal to zero and the switch resumes to an open circuit. In practice, V must be well below this limit to sample the input within a small enough time window. Single MOS switches are not favored therefore. For instance, in logic circuits, where the logic 1s and 0s are set by the power supply and ground, respectively, the logic high signal may be substantially corrupted and the speed reduced. CMOS switches are preferred threfore to single MOS switches. A typical CMOS transmission switch is shown in Fig. 14.2.6. It consists of the parallel combination of an

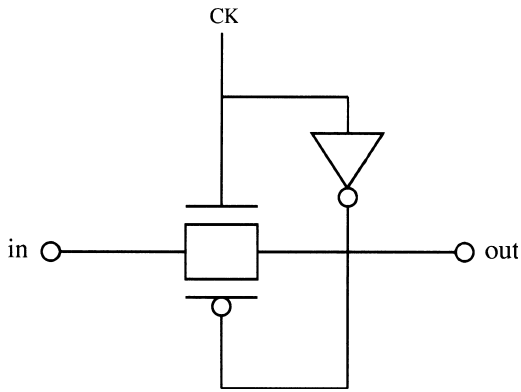

FIGURE 14.2.6 The complementary MOS switch.

N-MOS and a P-MOS transistor controlled by complementary logic signals. The idea is simply to counterbalance the decreasing conductance of the N-MOS transistor when the input goes from low to high by the increasing conductance of the P-MOS transistor. Thanks to the parallel combination, the series resistance is kept large and almost unchanged for any input voltage. The same holds true as long as the time is constant.

Dynamic Characteristics

MOS transistors ignore charge storage phenomena for they are unipolar devices. Their transient behavior is controlled by the parasitic capacitances associated with their gate, source, and drain. Source and drain are reverse-biased junctions, which exhibit parasitic capacitances with respect to the substrate. The gate capacitance relates to the inversion layer and to the regions overlaping source and drain. These capacitances control the dynamic behavior of the switch in conjunction with the parasitics associated to the elements connected to the MOS transistor terminals.

What happens with the inversion layer charge when the transistor is switched off is considered hereafter. Since charge cannot simply vanish, it must go somewhere, either to the source or drain or to both. This introduces generally a short spike in memoryless circuits that does not affect the performances significantly except at high frequency. In circuits that exhibit memory, like in the MOS sampling network discussed earlier, the impact is more serious. The part of the inversion layer charge left on the capacitive terminal is "integrated," which leads to a DC offset.

The charge partition problem in memory circuits is illustrated by the simple circuit shown in Fig 14.2.7, which consists of a MOS switch between two capacitors C_1 and C_2. We start from the situation where the voltages V_1 and V_2 across the capacitors are equal and no current is flowing through the transistor, supposed to be "on." As soon as the gate voltage starts to decrease, the charge in the inversion layer tends to divide equally between MOS terminals for these are at the same potential. If the capacitors are not identical, a voltage difference starts to build up as soon as charge is being transferred. This causes current to flow in the MOS transistor,

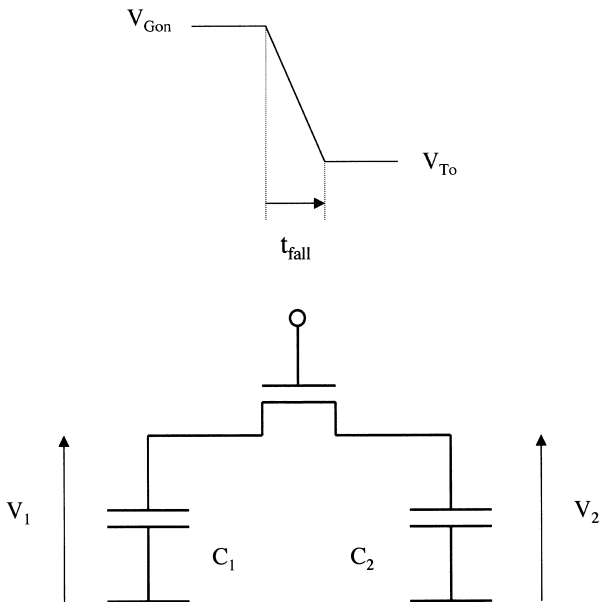

FIGURE 14.2.7 The inversion layer charge divides between C_1 and C_2 after cutoff.

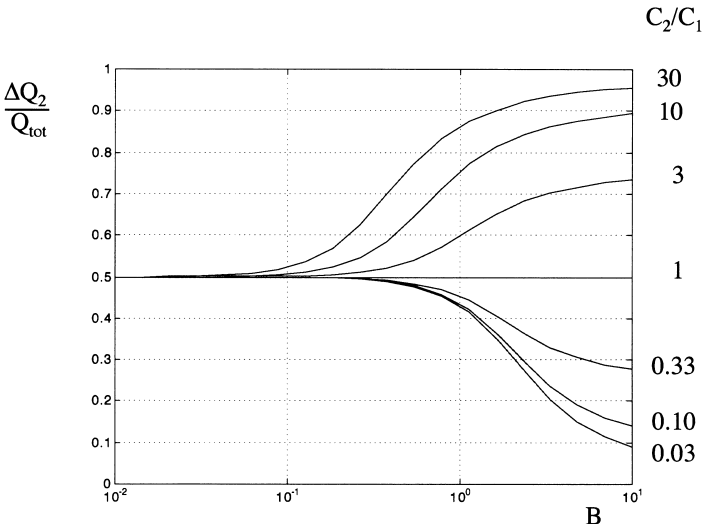

FIGURE 14.2.8 Fraction of the inversion layer charge left in C_2 after cutoff vs. the parameter B defined under Eq. (14).

which tend to reduce the voltage difference between the capacitors. This re-equilibration mechanism holds on as long as the gate voltage exceeds the effective threshold voltage although it is getting weaker and weaker as the transistor is nearing cut-off. When finally the MOS transistor is cut-off, a nonzero voltage difference is left over, which may be assimilated to an offset. The size of this offset is a function of several factors including the gate slewing-rate. It is obvious that an abrupt gate voltage step, which does not leave time for re-equilibration, will split the inversion layer charge equally between the two capacitors, whereas slow cut-off will tend to keep the voltages across the capacitors more alike.

The problem is addressed in Ref. 19. The fraction of the total inversion layer charge that is stored in capacitor C_2 versus the parameter B defined below is illustrated in Fig. 14.2.8:

$$B = (V_{\text{Gon}} - V_T) \cdot \sqrt{\frac{\beta}{aC_2}} \tag{14}$$

V_{Gon} is the gate voltage prior to switching, V_T the effective threshold voltage of the MOS transistor equal to $(V_{T0} + \lambda V_{\text{in}})$, β the well-known factor $\mu C_{\text{ox}} W/L$ and a the gate voltage slewing rate defined as $(V_{\text{Gon}} - V_{T0})$ divided by the fall time. Notice that fast switching yields small values of B, whereas long switching times lead to large values of B. When B is small the inversion layer charge divides equally. Voltage equalization tends to prevail when B is large, as can be found from the large differences experienced once the ratio C_2/C_1 departs from one. Let us consider, for instance, a MOS transistor with a β equal to 10^{-4} A/V^2, a gate capacitance C_G of 0.1 pF, V_{Gon} and V_{T0}, respectively, equal to 5 and 0.7 V, a large capacitor C_1 to mimic a voltage generator and a load capacitance C_2 equal to 1 pF. For fall times between 1 ps and 1 ns, the factor B varies from 0.021 until 0.626. The offset voltage is large and varies little from 215 to only 200 mV since re-equilibration cannot take place in such a short time. A 10 ns fall time reduces the final offset to 125 mV, and 100 ns, 1 μs, and 10 μs fall times yield, respectively, 41, 13, and 4 mV offset. In any case, these are still large offsets in terms of analog signals. To get smaller offsets the switching times must be very long. Hence, switching noise cannot be avoided as such.

A second important problem is nonlinear distortion. Less time is needed to block the MOS transistor when the input voltage is close to V_{Gon} for the effective threshold voltage becomes quite large. The amount of charge stored in C_2 varies with the magnitude of the input signal and the offset is thus a nonlinear replica of the input.

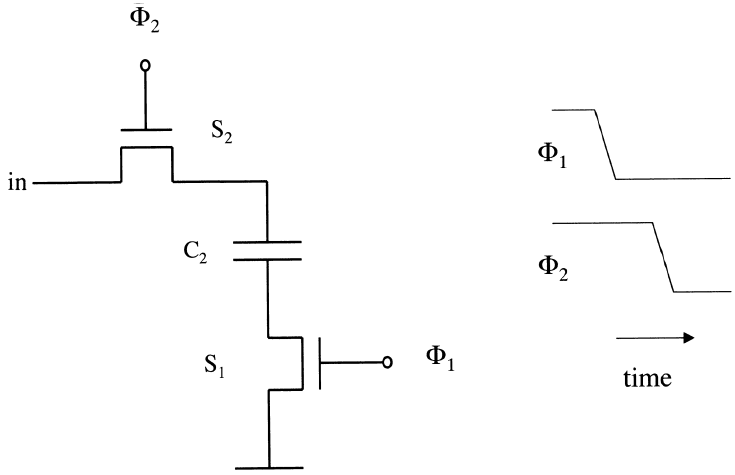

FIGURE 14.2.9 Switching noise nonlinearity can be lessened by means of the transistor S_2 in series with the storage capacitor.

This makes nonlinear distortion figures less than –70 dB hard to achieve unless a technique such as the one described hereafter is put to use.

In the circuit illustrated in Fig 14.2.9, the lower end of capacitor C_2 is tied to the ground by means of a second switch S_2. During the acquisition time, both S_2 and S_1 are conducting. Sampling occurs when the switch S_2 opens, shortly before S_1 opens. Suppose switching times of both transistors are short enough to avoid charge re-equilibration. When S_2 opens, the charge in the inversion layer divides equally between C_2 and ground. When S_1 opens, since C_2 is already open-ended, the signal dependent charge of S_1 has no other way out than to flow back to the generator. Thus, C_2 stores only charge from S_2, which is constant since the lower end of C_2 is tied always to ground. In fact, one exchanges a signal-dependent offset against a constant offset, which does not impair linearity. This offset moreover can be compensated easily by taking advantage of differential architectures that turn a constant offset into a common mode signal, which is ignored further on.

TRANSISTOR SWITCHES OTHER THAN LOGIC GATES

Transistor switches are extensively used in applications other than logic gates, covering a wide variety of both digital and analog applications. A typical illustration is the circuit converting the frequency of a signal into a proportional current, the so-called *diode pump*. This circuit (Fig. 14.2.10) consists of a capacitor C, two diodes D_1 and D_2, and a switch formed by a transistor T_1 and a resistor R. The transistor is assumed to be driven periodically by a square-wave source, alternatively on and off. When T_1 is blocked, the capacitor C charges through the diode D_1, while D_2 has no effect. As soon as the voltage across C has reached its steady-state value E_{cc}, T_1 may be turned on abruptly. The voltage with respect to ground at point A becomes negative, and D_1 is blocked, while D_2 is forward-biased. The capacitor thus discharges itself in the load (in Fig. 14.2.10 an ammeter, but it could be any other circuit element that does not exhibit storage), allowing V_A to reach 0 V before T_1 is again turned

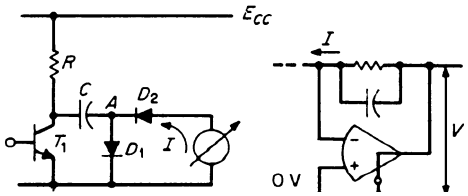

FIGURE 14.2.10 Diode-pump circuit.

off. The charge fed to the load thus amounts to CE_{cc} coulombs. If we suppose that this process is repeated periodically, the average current in the load is given by

$$I = fCE_{cc} \tag{15}$$

where f represents the switching repetition rate.

The diode-pump circuit provides a pulsed current whose average value is proportional to the frequency of the square-wave generator controlling the switching transistor T_1. The proportionality would of course be lost if the load exhibited storage, e.g., if the load were a parallel combination of a resistor and a capacitor in order to obtain the average current. Using an operational amplifier, as shown in the right side of Fig. 14.2.10, circumvents the problem.

The requirements on the switching transistor in this application are different and in many respects more stringent than for logic gates. The transistor in a logic circuit provides a way of defining two well-distinguished states, logic 1 and 0. Nothing further is required whether these states approach an actual short circuit or an open circuit. In the diode-pump circuit, however, the actual switching characteristics are important, since the residual voltage drop across the saturated transistor of Fig. 14.2.10 influences the charge transfer from C to the load, thereby also introducing unwanted temperature sensitivity. The main difference lies in the fact that while T_1 is operated as a logical element, the purpose of the circuit actually is to deliver an analog signal.

There are many other examples where the characteristics of switching transistors influence the accuracy of given circuits or instruments. An even more critical problem pertains to amplitude gating, since this class of applications requires switches which correctly transfer analog signals without introducing too much distortion. Furthermore, positive and negative signals must be transmitted equally well, and noise introduced by the gating signals must be minimized.

Analog gating. A typical high-frequency gating network for analog signals is shown in Fig. 14.2.11. Gating is performed by means of the diode bridge in the center. All remaining circuitry controls the on-off switching.

In order to transmit the analog signal, the current sources Qbar and Q must, respectively, be on and off. The current $2I$ from transistor T_1 is split into two equal components, one that flows through T_3, the other that flows vertically through T_4 and the bridge. The second forward biases all the diodes. Current, moreover, is injected horizontally from the signal source, left, to the output terminal, right. Those in- and out-currents representing the analog signal are equal since the sum of all currents injected in the bridge must necessarily be zero and the vertical current components though the bridge are balanced by the network. Voltage drops across the diodes are supposed to compensate each other.

When the path between source and load must be interrupted, the current sources Q and Qbar take opposite states. No current then flows through the bridge and the extra-currents supplied by T_2 and T_3 are diverted, respectively, through T_6 and T_5. The two vertical nodes of the bridge are now connected to low impedance nodes so that the equivalent high-frequency network between in- and output terminals consist actually of two branches, each with two small parasitic capacitances representing series reverse-biased diodes short circuited in their middle to ground. This ensures an excellent separation between in- and output terminals making this type of gating network well suited for the sampling of high-frequency signals, like those used in sampling oscilloscopes.

Field-effect transistors also are extensively used to perform analog gating. A typical application is switched-capacitor filters. Figure 14.2.12*a* illustrates an elementary switched-capacitor network. In this circuit, the capacitor C is connected alternatively between the two terminals so that a charge $C(V_1 - V_2)$ is transferred at each cycle. Hence, if the repetition rate is f, the switched-capacitor network allows an average direct-current $C(V_1 - V_2)f$ to flow from one terminal to the other. It is thus equivalent to a resistor whose value is $1/Cf$.

If another capacitor C_o is connected at the output port, an elementary sampled-data RC circuit is built with a time constant equal to C_o/Cf. This time constant depends only on the clock frequency f and on the ratio of two capacitors. Hence, relatively large time constants can be achieved with good accuracy using very small capacitors and MOS transistor switches. In practice, MOS capacitors of a few picofarads match better than 0.1 percent. In addition, a slight modification (see Fig. 14.2.12*b*) of the circuit avoids the stray capacitance, which would otherwise affect the accuracy adversely. Hence, fully integrated switched-capacitor RC active filters can be designed to tight specifications, e.g., for telephone applications.

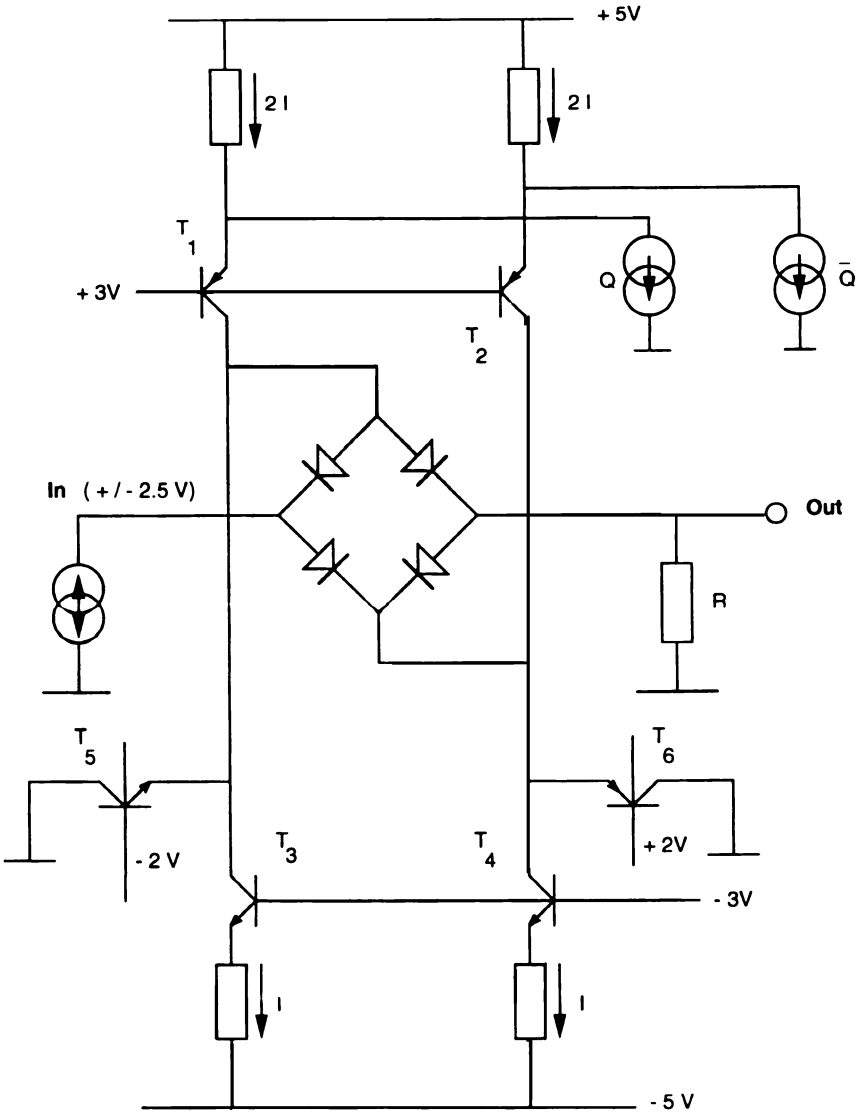

FIGURE 14.2.11 High-frequency gating network.

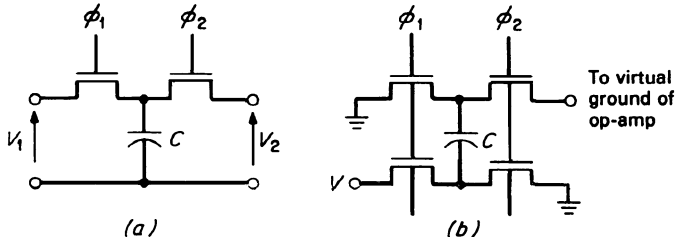

FIGURE 14.2.12 (*a*) Switched capacitor resistor; (*b*) the circuit is not affected by stray capacitances.