Switching Transistors

I. M. ROSS
NONMEMBER AIEE

Transistors are being used more and more frequently as components of computing systems. This increasing popularity can be attributed to their small size, high efficiency, and potential reliability. The size and efficiency of transistors are sufficiently well known to require no further discussion here. On the reliability question, it is pertinent to note that transistors have been in operation in some telephone applications for several years with failure rates of about 0.05% per thousand hours. This figure compares favorably with that for the best vacuum tubes and there is every confidence that newer devices will exhibit much higher reliability.

The first part of this paper is devoted to a discussion of the electrical characteristics that make conventional transistors (n-p-n and p-n-p) attractive in computing applications. The second part deals with a family of 4-region (p-n-p-n) devices which are now under development and in some cases in early production. These devices exhibit a bistable characteristic and their use may lead to a considerable simplification of computer circuitry.

Conventional Transistor Switches

A common form of switching transistor is shown in Fig. 1. Because such a device is frequently made by alloying techniques, it will be referred to it as an "alloy-type" transistor. One essential feature of such a structure is that the emitter and collector regions are metallic and hence, do not introduce appreciable series resistance in the emitter and collector leads. A further feature is that the emitter and collector junctions are opposite one another and are nearly the same size, the collector usually being somewhat larger than the emitter. Thus, the device is close to being symmetric.

Because of the lack of series resistance in the collector and emitter leads and also because of the symmetry of the device, this transistor is probably the most suitable for switching applications. It is further the most amenable to accurate analysis.

The simplest form of switching circuit using an n-p-n transistor is shown in Fig. 2. If the base current is very small or even negative, the current through the load resistor is of the order of the leakage current across the reverse biased collector junction and is micromperes or less. Thus, effectively all the supply voltage appears across the transistor which therefore acts like an open switch. When a current $I_b$ flows into the base lead, a current $I_b/1-\alpha$ will flow in the collector circuit. The collector voltage approaches zero as the collector current $I_b/1-\alpha$ approaches $V/R_L$. Effectively, all the supply voltage then appears across the load resistor and the transistor behaves like a closed switch. If the base current is increased beyond this critical value, the current that would be collected by a reverse biased collector becomes greater than the current that can be supplied through the load resistance from the voltage source. In fact, the collector junction becomes forward biased in order to reject or reinject some of the current and maintain the collector current close to the limiting value determined by the external circuit. Thus, for large values of base current, both the emitter and collector junctions are in the forward biased direction and the emitter to collector voltage, which is the difference of the two forward biases, becomes very small, as small as a few millivolts. This condition of operation with both junctions forward biased and the collector current saturated is referred to as the saturated condition. It should be noted that in the saturated condition, the emitter to collector voltage is less than the emitter to base voltage by the magnitude of the collector forward bias. Thus, if the collector of a saturated transistor is directly connected to the base of a second transistor the second transistor will be maintained in its off condition. The use of this mode of operation can lead to very simple switching circuitry.

A family of collector characteristics with base current as the parameter is shown in Fig. 3 for a silicon alloy-type transistor. For d-c operation, the most important parameters are:

1. The leakage current in the open or off condition. This depends on the dimensions of the device, the method of fabrication and most critically, on the material from which the device is made. For germanium transistors, the leakage current is of the order of microamperes at room temperature and doubles approximately every eight degrees Centigrade with increasing temperature. For silicon, the leakage currents are of the order of millimicroamperes and the temperature dependence is somewhat smaller than in germanium. For this reason, silicon transistors are preferred over germanium in many switching applications.

2. The "breakdown" voltage in the off condition. As the applied voltage approaches this value, the leakage current increases rapidly and the device no longer represents an open switch. Transistors can be designed to have breakdown voltages as high as 50 to 100 volts.

 Ross—Switching Transistors
The voltage in the on condition which for alloy-type transistors can be as low as tens of millivolts.

4. The current amplification factor $\alpha$, which determines the current gain, that is, the ratio of control to controlled current. In typical transistors, $\alpha$ is in the neighborhood of 0.9 to 0.99, giving a current gain in the circuit of Fig. 2 between 10 and 100.

Of prime importance are the times required to switch the transistor between the off and on conditions. These switching times depend not only on the parameters of the transistor but also on the particular circuit in which it is used. Unfortunately, no really satisfactory figures of merit have been devised to characterize the transient behavior of switching transistors. The discussion here is limited to a qualitative description of the transient effects and their dependence on transistor parameters.

Fig. 4 shows the transients which occur when a large signal pulse is applied to the base lead of the transistor in Fig. 2. The collector voltage starts out along an exponential fall determined by the amount of drive, the current gain and the transit time for carriers from the emitter to the collector junction. This transit time, which is inversely proportional to the frequency cut-off, $f_m$, of the transistor, is determined primarily by the base thickness, decreasing as the square of the base thickness. Thus, the thinner the base region the shorter the turn-on time $t_o$. In the saturated condition, both the emitter and collector junctions are forward biased injecting charge carriers into the base region and, as a result, the amount of charge stored in the base during the saturated condition is greater than in the unsaturated condition. This additional stored charge is roughly proportional to the lifetime, $\tau$, for minority carriers in the base and to the excess of base current over that necessary to drive the transistor into the saturated condition. When the base drive is removed, this additional stored charge must be removed before the collector voltage can start to rise. There is thus a storage time $t_i$, between the removal of base drive and the start of increase of voltage at the collector. The storage time $t_i$ depends on the amount of stored charge and the rate at which the charge can be removed. If the device is turned off by open-circuiting the base, the stored charge will decay with essentially the characteristic lifetime $\tau$. If however, the device is turned off by applying reverse bias across the emitter-base junction, the stored charge will also be drawn out at the emitter contact and, in the limit, can be removed in a time close to the transit time across the base. Thus, the storage time, $t_i$, depends on the degree to which the device was driven into saturation, the lifetime in the base region, the transit time across the base and the nature of the turn-off signal applied to the base contact. At the termination of the storage period the collector voltage begins to rise exponentially at a rate which is again determined by the frequency response of the device.

The device parameter common to the three delay times is thus the frequency response of the transistor which is closely related to the width of the base region. Since speed is usually the most important characteristic in a switching circuit, the designer is frequently willing to trade increased speed for a decrease in some other performance parameter. The diffused-base transistor is a typical example of such a compromise. Using this structure, the device designer can achieve thinner base regions and hence, increased frequency response but usually at the expense of several ohms of series resistance appearing in the collector lead.

It should be emphasized again that the delay times observed in a transistor switching circuit depend not only on the properties of the transistor but also on the particular circuit used. Transistors are now available which, in practical switching circuits, yield turn-on, storage, and turn-off times each in the range of 10 to 100 millimicroseconds. Hence, switching rates of 10 to 100 megacycles...
are practical. There is no doubt that the trend in manufacture towards higher frequency transistors will continue over the next few years and hence, devices should be available which will yield switching rates considerably higher than 100 megacycuses.

**Four Region Devices**

Fig. 5(A) shows a schematic diagram of a p-n-p-n diode biased so that current flows from left to right. This direction of current tends to make the outer two junctions forward biased and the middle junction reverse biased. The device is designed so that the outer regions are good emitters into the middle regions and that there is substantial transmission of carriers across the middle regions. Thus, we can look upon the device as being a combination of two transistors, as shown in Fig. 5(B), there being two separate emitters and a common collector. Defining the alphas of the p-n-p and n-p-n transistors as being \( \alpha_p \) and \( \alpha_n \) respectively, the current across the middle reverse-biased junction is the sum of the following three components: a leakage current \( I_L \) due to the bias across the middle junction itself, a hole current \( I_{h} \) due to injection from the p-type emitter and an electron current \( I_{e} \) due to injection from the n-type emitter. Since the sum of these three components must equal the total current \( I \) in the device,

\[
I = I_L + I_h + I_e
\]

which gives for \( I \)

\[
I = I_L / [1 - (\alpha_p + \alpha_n)]
\]

From equation 2 it is seen that, if the sum of alphas is less than unity, the current is a multiple of the leakage current, and therefore, the VI characteristic of the device is similar to that of a reverse biased p-n junction, curve 1 of Fig. 5(C). Now consider what happens if the sum of alphas is greater than unity (since each alpha can be as large as unity, the sum can easily be greater than unity). Equation 1 shows that for this case the injected current that would be collected if the middle junction were reversed biased is greater than the total current that flows. Just as in the case with the saturated transistor, the middle junction becomes forward biased to reject or reinject some of this current. Hence, for sum of alphas greater than unity, all three junctions are forward biased and the net voltage across the device is approximately equal to the forward bias across a single p-n junction. The VI characteristic is then similar to that of a forward biased diode, curve 2 of Fig. 5(C). Since, in silicon transistors, a increases with increasing current, it is possible to design a silicon p-n-p-n diode in which the sum of alphas is less than unity at low current and greater than unity at high current. The resulting VI characteristic is shown as curve 3 of Fig. 5(C) where \( I_b \) is the current at which the sum of alphas goes through unity. This characteristic is similar to that of a gas discharge tube and as with the gas tube an load line can be chosen to give two stable states of operation, one at low current and high impedance and the other at low voltage and low impedance. Unlike the gas tube, the sustain voltage, i.e., the voltage in the on condition, is very low, less than one volt. Hence, as a switch, a p-n-p-n diode can be much more efficient than the gas tube. Fig. 6 gives the characteristic of a typical silicon p-n-p-n device. This diode in the high impedance condition has a leakage current of a few millimicroamperes shunted by a capacitance of about 10 micromicrofarads. In the on condition the sustain voltage is about 0.8 volts and the dynamic impedance about 2 ohms.

The transient behavior of a 4-region diode is similar to that of the 3-region transistor being characterized by "turn-on," "storage," and "turn-off" times. The turn-on time is essentially determined by the transit times across the two bases. In developmental models, turn-on times as low as 10 microseconds have been observed. Storage time is again determined by the lifetime, the transit times in the bases and the extent to which the device was driven into the low impedance condition. Little information is yet available on the values of storage time achieved in this device; however, values comparable to those for transistors should be possible.

An essential feature of the p-n-p-n diode is that it reaches the low impedance condition when the current through the device reaches a critical value, that is, the value for the sum of alphas equal to unity. In the 2-terminal device, current is caused to increase by the diode going into the breakdown region. An alternate way in which the current can be made to increase is to put a contact onto one of the middle regions and forward bias the outer junction. It is then possible to switch from the high impedance to the low impedance condition without biasing up to the breakdown voltage. The family of characteristics obtained from an experimental 3-terminal device is shown in Fig. 7 and is reminiscent of that for a thyratron tube. However, unlike the thyratron, it is possible to turn off the device by pulling current out of the base contact.

There are two other useful ways in which the p-n-p-n diode can be switched. The first method makes use of the fact that semiconductors are photosensitive. If sufficient light is shone on a p-n-p-n diode in the off condition, the current can be increased beyond the switching current \( I_b \). The device can, therefore, be used as a photo switch. The second method involves the fact that there is a shunt capacitance associated with the reverse biased middle junction. Thus if a sharp rise in voltage is applied across the device, the displacement current necessary to charge the middle junction can turn on the device. Although this transient turn-on may find useful applications, it may also act as a speed limitation in other applications.

The bistable nature of the characteristics of p-n-p-n devices makes them attractive components for numerous computing applications. As illustrated in Fig. 5, the device is equivalent to two interconnected transistors and hence, its use can result in a considerable decrease in the number of components necessary to perform a given function. It is also possible that in the future, complete functional circuits can be built into one semiconductor device. For example, using p-n-p-n material, experimental models of 4-stage counters have been made, the transfer from stage to stage occurring totally within the semiconductor. These devices, which are electrically analogous to gas stepping tubes yet the size of a normal transistor, have operated at counting rates up to 1 million per second.

**References**

Special-Purpose Tubes for Computer Applications

Saul Kuchinsky
Nonmember AIEE

Special-purpose tubes have received acceptance as active elements in the performance of many fundamental electronic functions. Three general classifications of such tubes are of special interest in computer applications: beam switching tubes, in-line alphanumerical indicators (Nixie), and cathode ray readout tubes (Charactron, Tyantron).

The beam switching tube is a high-vacuum electronic distributor and multioutput device which uniquely performs such functions as counting, timing, programming, sampling (telemetering), dividing, coding, matrixing (memory addressing), and binary decoding. It differs from other active elements in that a single cathode controls an electron beam to any one of ten constant current output positions, each of which has provisions for bistable beam locking and high impedance switching.

Gas-filled in-line alphanumerical indicators, commonly called “Nixies”, represent another type of special purpose tube with characteristics not readily available by other means. Elements in the form of numerals, letters, or special characters may be selected by applying voltage to appear as a glowing cathode in a simple gas discharge. The phenomenon whereby the visual glow discharge is larger than the actual cathode permits the display of all in-line characters in a common viewing area. The device is an unusually efficient electronic to visual converter since almost all of the electronic energy is concentrated in a visual glow of relatively narrow optical bandwidth. The eye acts as a natural filter in distinguishing this glow in high ambient light. Both decimal and biquinary-type readouts are described.

Charactrons and Tyantrons represent still another type of important special purpose tubes. These cathode ray tube types are now described herein are performing increasingly important computer functions where high-speed printout and readout characteristics are primary requirements.

Beam Switching Tubes

The beam switching tube is a 10-position high-vacuum constant current distributor (Fig. 1). It consists of ten identical “arrays” located radially about a central cathode (Fig. 2). Each array comprises, 1. a spade which automatically forms and locks the electron beam, 2. a target output which makes the beam current available with constant current characteristics, and 3. a high impedance switching grid which serves to switch the beam from target to target. A small cylindrical magnet is permanently attached to the glass envelope to provide a magnetic field which, in conjunction with an applied electric field, comprise the crossed fields necessary for the operation of this tube.

The tube may be made to operate in almost any conceivable distribution or switching mode such as 1. the tube may be in a clear or cut-off condition, 2. an electron beam may be formed in any one of the ten positions, 3. the electron beam may be switched sequentially, 4. the electron beam may be switched at random from any one position to any other position, and 5. the electron beam may be switched and cleared (cut itself off).

The impedances of the three electrodes are such that building block versatility exists. A distributor or switch of any number of positions may be obtained, since there are simple methods of arranging circuit parameters and interconnections between positions in the same envelope or between separate tubes. The same versatility permits accepting binary inputs directly for converting to decimal or other codes.

Fig. 3 indicates a typical test and operating circuit for the type 6700 Tube. The operation of the tube can best be understood by studying the characteristics of the three basic elements comprising each array.

Spade Characteristics

Clear Condition

The spade electrodes directly affect the magnitude and shape of the electron beam in the area between the cathode and the spades. The tube will always be in the cut-off state when power is first applied if there are no provisions for beam forming. The spades are commonly connected to their supply voltage through individual series load resistors. When all of the spades are at B+ potential, the tube is equivalent to a magnetron diode in the cut-off condition, since the magnetic field prevents any electrons from reaching the outer arrays.

Beam Formation—Static From Cut-Off

The beam may be formed in any one of its ten “on” positions by sufficiently lowering the potential of the respective spade with either a d-c voltage or a high-speed pulse. Each spade has a negative resistance characteristic due to the crossed electric and magnetic fields. The resultant bistable states are shown in Fig. 4 by the solid “static” curve intersected by the series spade resistor load line. When the spade potential is lowered to approximately 60% of the spade B+, the negative characteristic will provide automatic lock-in at a point near zero cathode potential. Thus, the one spade which forms and locks the beam is near cathode potential, while the remaining spades remain in the cut-off state.