

Physics of digital devices

Robert W. Keyes

IBM Research Division, Yorktown Heights, New York 10598

Large computing machines have been built from relays, vacuum tubes, and transistors. Strenuous efforts to develop computing technology based on other devices have failed. The relay, vacuum tube, and transistor are three-terminal devices, and the essential physical factors that account for their success are described and used to discuss the reasons for the failure of other devices.

CONTENTS

I. Introduction	279
II. Large Systems	280
III. Digital Representation	280
IV. The Transistor	281
V. Variability of Devices	282
VI. Other Devices	282
A. Bistability	283
B. The Esaki diode	285
C. The Josephson cryotron	286
VII. Conclusions	287
References	287

I. INTRODUCTION

Figure 1 shows a part of the Automatic Sequence Controlled Calculator (ASCC) built by IBM and presented to Harvard University in 1944 (Harvard, 1946). The rectangles lined up in rows and columns are relays, electrically

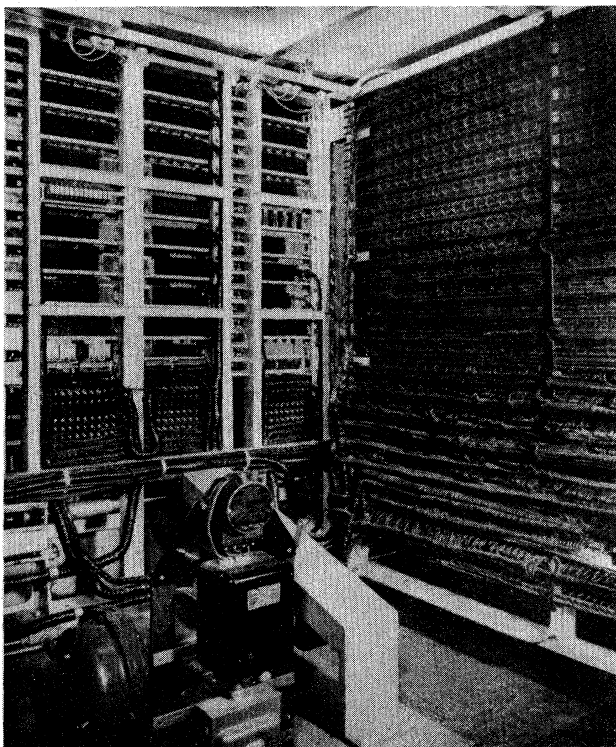


FIG. 1. View of the Automatic Sequence Controlled Calculator (from Harvard, 1946).

controlled mechanical switches. A closer view of the switching elements of the machine is presented in Fig. 2. The ASCC and other early relay-based calculating machines demonstrated the power and usefulness of automatic calculation and stimulated a desire for faster, more powerful systems. A search for better basic component devices than those shown in Fig. 2—components that were smaller, faster in operation, lower in cost, less demanding of power, and more reliable—was touched off. Great progress has been made in the intervening four decades, but the quest for devices that would be improved in terms of all of those properties continues to the present day.

The history of the search for better devices for computing systems is punctuated by many ingenious proposals. Most did not propagate far beyond the sphere of influence of their inventors. A few received wide attention and became the focus of well-funded development efforts. Besides the relay, however, only two turned out to possess the qualities required of devices that are to be used to construct the logic in large computational systems: the vacuum tube and the transistor.

The many attempts and rare success stories in the history of the pursuit of still better logic devices lead one to ask what features distinguish the successful devices from those that attained only oblivion. More particularly, are there characteristics that will enable a good device to be recognized without investment in a substantial development program? This paper suggests that the answer to the question is yes.

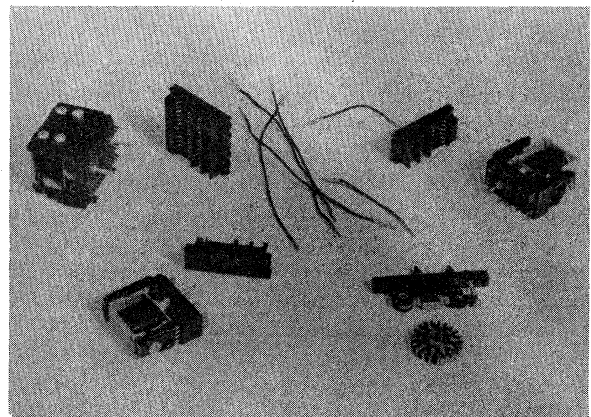


FIG. 2. Components used in the Automatic Sequence Controlled Calculator (from Harvard, 1946).

II. LARGE SYSTEMS

In order to explain this answer and the identification of the attributes necessary in a device, some characteristics of large logic systems must be identified and discussed in a general way. Computers are large in the sense that they contain many devices, tens of thousands to several millions. These large numbers severely constrain the physical representation of information in systems and the properties of devices used to implement them. Many types and lengths of problems can be handled by computers; short and very long calculations can be performed. The course of a calculation may be conditioned by frequent tests, branches, and memory references. All of this involves great "depth," in the sense that information is handled over and over again (von Neumann, 1958). A large amount of communication among the many devices in a computer is entailed. The output of one device is rapidly input to another, and so on through hundreds or thousands of steps. There are frequent opportunities for a signal to be altered in its passage from one device to another, through effects known as attenuation, dispersion, diffraction, or crosstalk. Even a 1% distortion per step could lead to complete loss of information in less than a hundred steps. The loss of information through the accumulation of such distortions in hundreds or thousands of steps is avoided by setting the output of a device to one of a set of standard values after each step. The standard values are sufficiently widely separated that they can be distinguished easily at the next stage. Accumulation of the losses of fidelity is thereby prevented.

The output of a device may be required as input by several other devices. The transmission of a signal to a multiplicity of destinations is known as fan-out; devices for computers must be able to provide fan-out. Furthermore, it is desired that a computation in a machine proceed in one direction, from inputs to final results. Each device should operate only on its inputs and not be sensitive to the actions or status of the recipients of its outputs. This property is known as input-output, or I/O, isolation and is required in computer devices.

III. DIGITAL REPRESENTATION

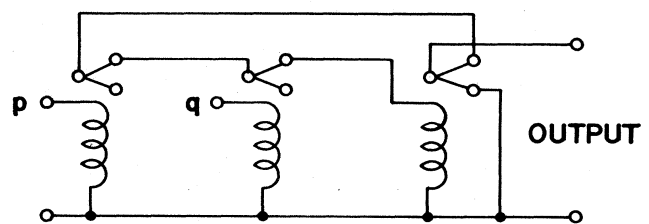
The representation of information by a finite set of standard values is known as digital representation. Digital representation is, of course, familiar. Ten digits are used to represent numbers in most of the world. Computing machinery uses only two digits, the binary system. The use of binary notation is especially well suited when information is to be represented as the value of some physical quantity. The construction of a switch that is either open or closed is easy and permits an electrical current or voltage to either appear or not appear at a certain point. Many physical phenomena are inherently binary or nearly so: a bar of iron can be magnetized in either of two directions; a site for a hole in a card is transparent or opaque; certain metals may be normal or

superconducting. Often, too, a binary choice is all that is needed to carry desired information: a red light and a green light convey the status of a road intersection; a pilot light signals that a stove is turned on; a single bit (the name for the amount of information represented by a binary digit) tells the computer whether a password is out of date.

Digital representation also allows arbitrary accuracy simply by using many digits. The value of pi is known to millions of decimal digits, for example.

The features of binary logic are illustrated in Fig. 3 with a relay circuit. In terms of its electrical characteristics, the relay is a nearly ideal computer device. The figure shows a NOR circuit and its truth table, which describes its logical function. Terminals A and B are input terminals, and O is an output. Power is supplied via the +V and ground terminals. In terms of the table, 1 means the presence of voltage. An input voltage applied to A or B causes relay R to be energized, grounding output O. If neither A nor B receives a voltage input, then O is connected to +V; a 1 is transmitted.

Note those aspects that make the relay eminently suitable for operation in a large computing system. The values of the output are standardized by connections to +V and ground. These potentials are distributed everywhere in the system; a zero and a 1 are the same throughout. The information is digital, and large distortions of the signals during transmission are endurable. Figure 4 shows how the conductance between the relay contacts depends on the input voltage applied to the coil. Switching will take place even if there are substantial deviations from the nominal signal values. The existence of these ranges of input signals over which the correct outputs are obtained are known as "noise margins." The noise margins also provide tolerance of variability in the relay characteristics. The voltage at which the contacts close, V_T in Fig. 4, does not need to be controlled pre-



p	q	p NOR q
0	0	1
0	1	0
1	0	0
1	1	0

FIG. 3. Relay NOR circuit and the NOR truth table.

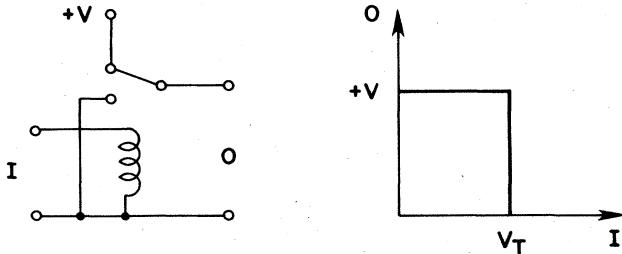


FIG. 4. Input-output I/O characteristic of a relay.

cisely. V_T can vary through a substantial range and the relay will operate correctly, with some sacrifice of the variability permitted in the signals.

There is excellent I/O isolation in Fig. 3; the closing of the contacts produces almost no signal in the coil that can be transmitted back along the input leads. Fan-out is easily obtained; the current that can be carried by the contacts is available to drive other stages and can be much larger than the current needed to actuate a relay.

Standardization of signals and fan-out require that an electrical device control voltages and currents larger than those needed to operate it, in other words, that a device have both current and voltage gain. High gain is necessary in order that the transition at V_T occupy a small part of the signal swing and allow the high noise margins. Gain is essential to digital devices.

Although the relay possesses the qualities needed in a digital device in an exemplary way, its application is limited by other constraints imposed by large systems. The construction of large systems, containing from thousands to millions of circuits that perform functions resembling the NOR (Fig. 3), is only practical if the cost per circuit is very low. The power dissipation must be small or the problem of removing the heat produced would be insurmountable. Moreover, low failure rates are essential; the system will not function unless all components work properly. If the average failure rate of each element of a system of 100 000 elements is once in ten years, failures of the system will occur at an average rate of one per hour. Finally, it is desirable that components be physically small. Small size makes short distances between components possible, and, since signals only travel with a finite velocity, short distances between components enable a system to operate faster. Of course the objective of fast operation is also aided by devices that need only a short time to produce an output after receiving an input. Ways to improve upon the relay in all of these respects were found.

IV. THE TRANSISTOR

The vacuum tube actually entered computer technology almost simultaneously with the relay. Its principal advantage as compared to the relay was faster operation; it was completely electronic, eliminating the slow mechani-

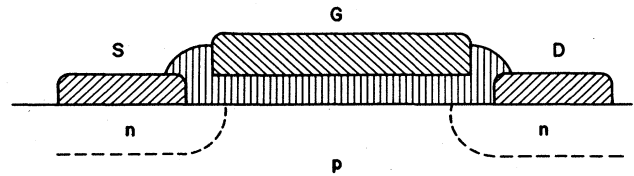


FIG. 5. Structure of an insulated gate field-effect transistor. A positive voltage applied to the gate G attracts electrons to the surface of the semiconducting substrate, establishing a connection between the source and the drain.

cal response involved in relay operation. The transistor, after its invention, quickly displaced both the relay and the tube, however. The transistor is also completely electronic, and its great potential for small size, low power, high reliability, and fast operation was quickly recognized. The development of the integrated circuit dramatically reduced costs. Thus transistors have been the only device used in the logic circuitry of computers for three decades.

Figure 5 shows a cross section of a field-effect transistor, an FET. Its operation depends on the application of a positive voltage to the gate electrode, which attracts electrons to the surface between the insulator and the semiconductor, thereby establishing a connection between the source and drain electrodes. When the connection is established the transistor is said to be "on," otherwise "off." The FET can be used to build a NOR circuit, as shown in Fig. 6(a) (Wallmark and Carlstedt, 1974). Again, $+V$ and ground potentials are distributed through the system. The circuit and device emulate the relay of Figs. 3 and 4; the dependence of output on input in the FET circuit is presented in Fig. 6(b). The connection to $+V$ is made through a transistor connected to act as a nonlinear resistor. The value of the resistor is much greater than the resistance of the FET in its turned-on state, so that the output voltage is close to ground potential when a positive input is applied.

The high gain and large nonlinearity of the response characteristic in Fig. 6(b) are again essential to the functioning of the circuit in a system environment. As in the case of the relay, the response provides large noise margins and tolerance of variability at the point at which the FET becomes conductive, its threshold voltage. The effect of the signal voltage on the FET is determined by its value relative to the power supply voltages that are distributed through the system. The input to the FET is capacitive; electric charge must be fed through the inputs to the gate electrode, Fig. 5, to turn it on. There is excellent I/O isolation. Fan-out is achieved by allowing current to flow until sufficient charge has been supplied to all of the following inputs. Unlike some devices, which will be mentioned below, the circuit can be switched both on and off by input signals in comparable amounts of time. Notice also that large noise margins separated by a region of high gain imply a very nonlinear

response. Nonlinearity is another essential feature of a digital device.

Similar features are found in logic circuits that use bipolar transistors: high gain and large nonlinearity that allow adequate noise margins, adaptability to large fan-out, and switching threshold determined by a comparison of a signal with standard voltages of the system (Kohonen, 1972).

Transistors have proved to be remarkably adaptable to computing systems. They seem incredibly susceptible to miniaturization and integration; all predictions of a limit or a slowing of the rate of changes in these characteristics have turned out to be pessimistic. The variety of types of transistors and the two polarities of charge allow many kinds of circuits to be built, so that transistors are the exclusive form of logic device used in all systems, from 4-bit microprocessors to the largest supercomputers. Transistors, of course, are also found in other applications, particularly, in computers, in memories, where several millions of transistors can be formed on one chip.

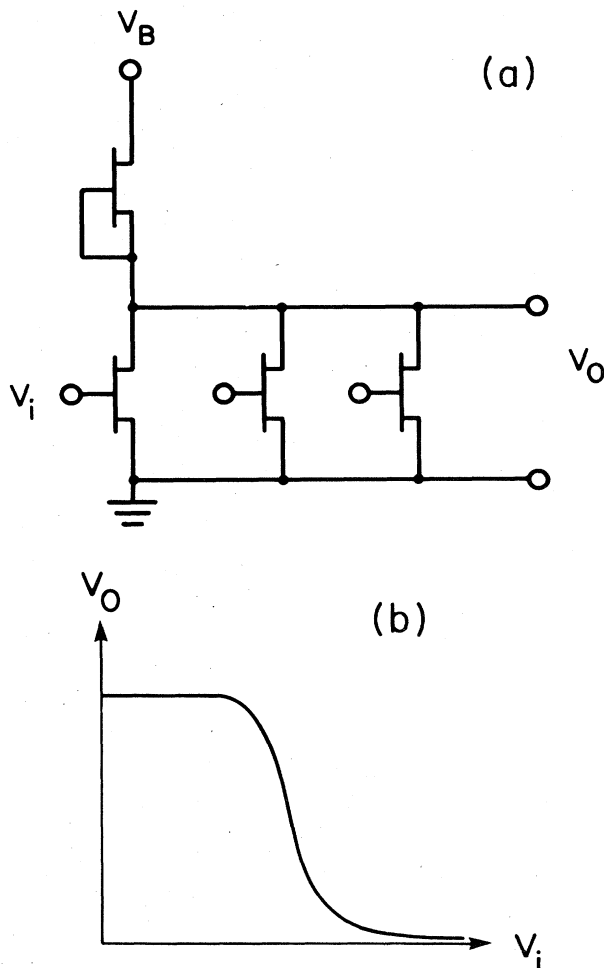


FIG. 6. Logic with field-effect transistors: (a) NOR circuit made from FET's; (b) transfer characteristic of the circuit of (a).

V. VARIABILITY OF DEVICES

Substantial noise margins are needed in systems, both because of the variability of signals, discussed earlier, and because of variability in the devices themselves. The environment in the computer is an important source of the latter. The energy that is used to operate devices is converted to heat, which must be caused to flow out of the system by temperature gradients. The temperature within the system therefore varies from point to point. It depends, too, on the temperature external to the system, the eventual destination of the heat currents. Further, the heat produced depends on the operations being performed, and varies from time to time. The differing temperatures within the system cause differences among the devices, as they affect such fundamental solid-state properties as energy gaps and the excitation of lattice vibrations.

In addition, mass production methods must be used to achieve the low costs that enable very large numbers of devices to be used. The low cost of semiconductor devices is attained by integration, the fabrication of many devices in a single solid body, which makes access to individual devices difficult or impossible. In any case, the objective of low cost precludes attention to adjusting or repairing individual devices. The limitations of the original manufacturing system in such matters as precision of dimensions and purity of materials remain with the devices and are another source of variability.

Finally, devices change with time. Such effects as corrosion and creep due to strains resulting from attachment of devices to some substrate are not completely avoidable. The use of devices in high-speed circuits subjects them to strong energy fluxes that enable or drive the motion of intrinsic defects, causing processes of degradation that include, but are not limited to, thermomigration and electromigration.

Noise margins must be sufficient to allow circuits to operate in the presence of all of these sources of variability.

VI. OTHER DEVICES

However, in spite of the long-continued success of transistors in computers and the absence of any evidence of slowing of progress in transistor technology, proposals for alternative devices have emerged regularly during the transistor era and have attracted interest and support. Indeed, some have been regarded as very promising alternatives to transistors for logic and have been the focus of substantial development programs. The development efforts have, however, turned out to be of no avail. Thus one is tempted to ask the following: what is the difference between those devices that have been successfully used to build large computers and those that have been unsuccessful in spite of the devotion of considerable quantities of resources to their development? In other words, are there criteria that can be used to evaluate the

potential of a device as a logic element without launching a massive development project?

In fact, the preceding text has explained the qualities necessary in a device intended for use in a computing system. The question to be answered is the following: does the device possess these qualities? The application of the criteria to the relay and the transistor has also been discussed, and the criteria will be applied below to unsuccessful devices for additional illustration of the points. First, however, it should be mentioned that the ideas involved are not really new; they were generally understood by von Neumann (1958) and Lo (1961) some time ago. The experiences of the years that have elapsed since these authors first explained the principles of digital computation have illustrated and amply confirmed their views.

A. Bistability

The proposed alternatives to the transistor have differed in a very fundamental way from the relay, vacuum tube, and transistor: they are two-terminal bistable devices or circuits. Bistability can be produced readily with a physical system that exhibits a negative differential electrical resistance. Negative resistance means that the slope of a plot of voltage versus current or of current versus voltage is negative. The physical origins of negative resistance are quite diverse, and it is not an infrequent occurrence. Similar effects are found in other physical relationships. The form of a bistable response is shown in Fig. 7(a). Quantity Y depends on X in such a way that there is a region in which Y decreases as X is increased. Now think of X being increased from zero. The value of Y is determined by the lowest branch of the curve of Fig. 7 until X reaches the value T , a threshold. When X exceeds T , Y can no longer follow the lowest branch; it must jump, or switch, to the topmost branch, which determines Y as X is increased above T .

Next, consider the sequence of events as X is decreased after having exceeded T . Now Y is determined by the upper branch of the curve until X reaches R , at which point it must switch to the lower branch of the curve. "Bistability" means that when X is between R and T , Y may have either of two values, depending on the preceding sequence of events, and it remains on one or the other segment of the characteristic curve until it is changed by passing X through R or T . The negative-slope part of the curve is never observed; in effect, the device has the characteristics shown in Fig. 7(b).

One may wonder how this bistable response can be used to implement logic functions. A two-input AND function, defined in Fig. 8, will be described as an example with the aid of Fig. 9. The value of X is formed by adding inputs in logic with bistability. A steady signal $X = B$, a bias, with B between R and T , is one component of X , and it maintains the device on the lower branch. Signals of magnitude S from other parts of the computer are added to X to perform the logic function. The signals

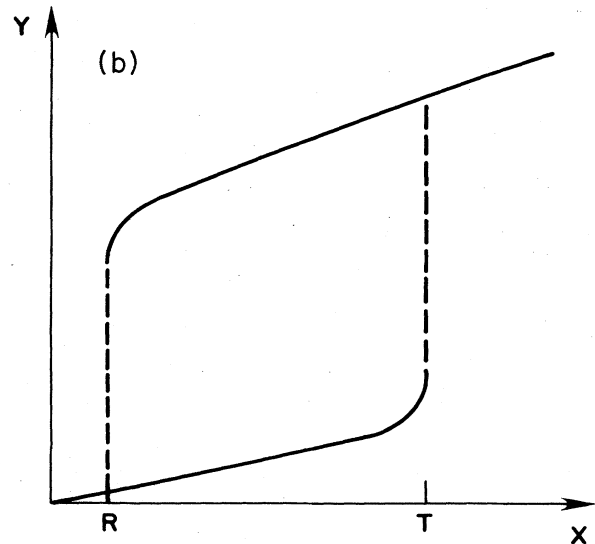
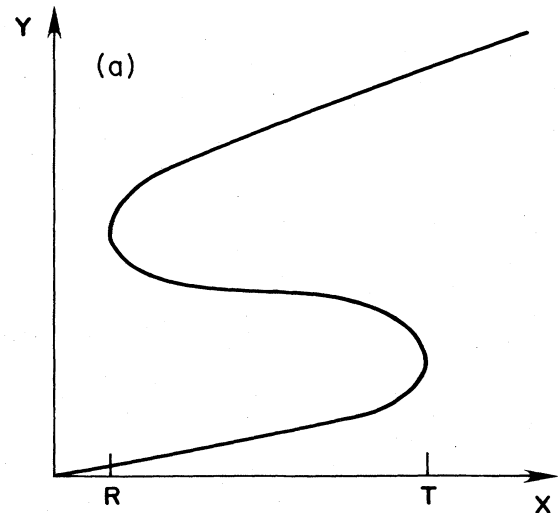


FIG. 7. Input-output characteristic of a bistable system: (a) characteristic of a bistable system; (b) response of (a) as it appears to the user.

represent the output of other logic stages and, ideally, have value zero or the design signal level S . Bias B is chosen so that $B + S$ is less than T , but $B + 2S$ is greater than T . Two signal inputs cause the device to switch to the upper branch; a measurement of the state of the device, even after the signals are removed and X is again equal to B , allows a determination of whether two

p	q	$p \text{ AND } q$
0	0	0
0	1	0
1	0	0
1	1	1

FIG. 8. AND function.

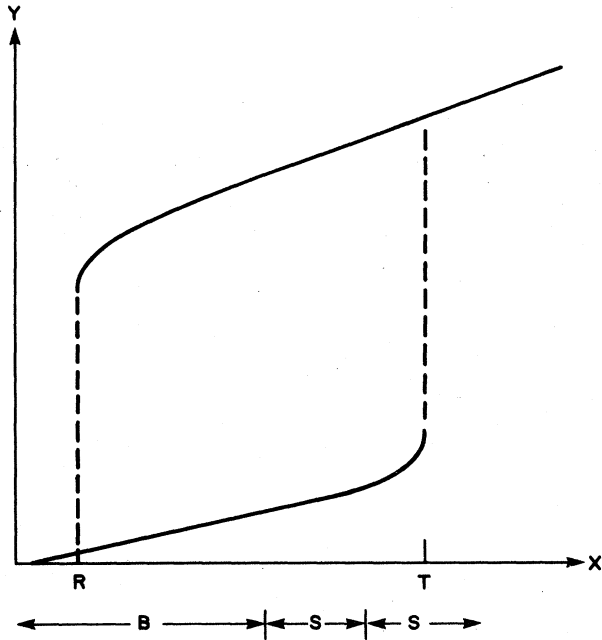


FIG. 9. Use of the response curve of Fig. 7 to perform the AND function. The bias input B maintains the device on the lower branch of the curve; the addition of two signals S causes switching to the upper branch.

nonzero signals arrived, or only one or none, which is the purpose of the AND, Fig. 8. The value of Y can be transmitted as output to the next stage. The device can then be reset to the lower branch by decreasing X below R and increasing it to B , readying it for the next operation.

It is not difficult to see that such bistable logic fails to satisfy the criteria developed above. The signal Y transmitted to constitute an input to a following stage is determined by the characteristic of the transmitting device, rather than by any standard. The lack of a reference point to establish the signal value reflects the variability of devices into the signals.

The bistable characteristic is unfavorable for logic in other ways. There is no operation of inversion, in which the presence of a signal at an input is indicated by the absence of a signal at the output. A separate resetting operation, in which the bias is reduced below R , is needed between logic operations to ensure that the device is on the lower branch of the characteristic when the inputs arrive. These features complicate the circuitry and operation of systems based on the device.

Another characteristic of logic with bistable devices, an intolerance of device variability, can be seen by writing the conditions that must be satisfied by the parameters of the system in the two-input AND:

$$B + S < T, \quad B + 2S > T .$$

The bias plus one signal input must not exceed the threshold, and the bias plus two signal inputs must exceed the threshold. These relations are illustrated

graphically in Fig. 10, in which the horizontal coordinate represents the design value of the bias, and the vertical axis the chosen value of the signal S . A choice of these variables is represented by a point in the figure. The fact that B must be between R and T means that the operating point must lie between the vertical lines passing through R and T on the horizontal axis. The condition $B + S < T$ means that the operating point lies below the diagonal line corresponding to $S + B = T$. Also, it must lie above the line representing $B + 2S = T$. The region available as an operating point, in which all three of these relations are satisfied, is indicated by the horizontal shading in Fig. 10.

If the bistable device is intended for use in a large system, the design considerations summarized by Fig. 10 must be extended to take account of the variability of devices. The values of S and T will not be known precisely, and the device must be made to work correctly in spite of the uncertainty in its characteristics. Figure 11 incorporates this additional aspect into the design of Fig. 10. To simplify the illustration, only the variability in the threshold T is considered.

Thus let T lie between T_1 and T_2 , lying on either side of a nominal threshold T_0 . Lines as in Fig. 10 are drawn to indicate the boundaries of the possible operating regions for $T = T_1$ and $T = T_2$. The allowed regions are indicated by different shadings for the two limiting cases. As the device must work in both of these extremes, the operating region must lie in the area common to both of the shadings in Fig. 11. It is seen that the region in which the operating point must lie is greatly contracted as compared to Fig. 10. Difficulty in ensuring that the signal lies within the narrow bounds permitted throughout a system can be anticipated, especially in

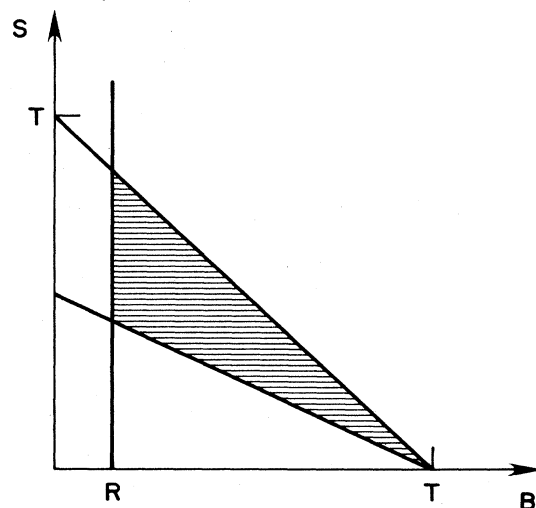


FIG. 10. Restrictions on the operating point of Fig. 9 in the B - S plane. The values of S and B must correspond to a point in the shaded region of the graph, which is bounded by the lines $B = R$, $B + S = T$, and $B + 2S = T$.

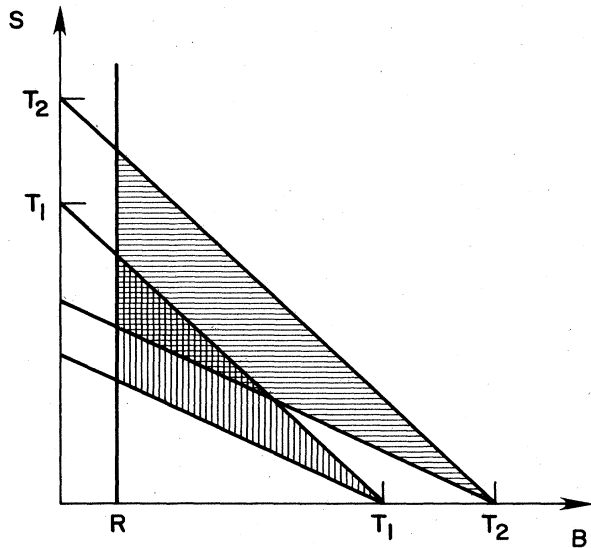


FIG. 11. Effect of variability of the threshold on the restrictions shown in Fig. 10. The two types of shading represent the allowed regions in which B and S must lie, as in Fig. 10, for the two extreme values of the threshold T_1 and T_2 . In order to ensure proper operation through the entire range of thresholds, the operating point must be in the region where the two shadings overlap.

view of the variability in S arising from both the unstandardized logic outputs and distortions of a signal during propagation from one point to another. Precise control of all parameters of a device is required. This need for tight control of the characteristics of devices has proved to be impossible to satisfy in systems of a great many components.

Note that the problem only appears in large assemblages of components. One or a few devices can be made to work very well in a laboratory environment where adjustments can be made to fit the characteristics of particular devices, temperatures can be carefully controlled, and high reliability of operation over a long time is not demanded. The limitations of the devices are not apparent until attempts to construct a system containing hundreds or thousands of devices that must all work simultaneously and communicate with one another are undertaken.

A limitation on a form of gain attributed to bistable devices is another consequence of the variability in characteristics. Control of an output much larger than the signal causing switching can be demonstrated by biasing the bistable device very close to the threshold, so that only a very small signal must be added to induce switching. The proposed source of gain requires that the threshold be known precisely, so that the bias can be adjusted properly. Thus much larger gains can be achieved with one device under controlled conditions in a laboratory than are available in systems of many devices, where

the bias must be chosen to be below the entire range of thresholds.

B. The Esaki diode

As an example of bistable logic, the Esaki or tunnel diode will be discussed (Gentile, 1962). The dependence of voltage across the tunnel diode on the current passing through it is shown in Fig. 12. The electrical circuit proposed to use it to perform the AND function is illustrated in Fig. 13. A current flows from the power supply, voltage V_B , through the resistor R_L and the diode TD to ground. The current is determined by the condition that the voltage drop across the diode, given by the curve of Fig. 12, plus the ohmic voltage across the resistor, must sum to V_B . This condition is satisfied at the points of intersection of the tunnel-diode voltage-current characteristic with the load line, the dotted line in Fig. 12. There are two stable intersections. Logic operations are performed with the circuit of Fig. 13 by starting with the diode in the low-voltage state. Then the signal currents are injected into the circuit through the resistors R_1 and R_2 . The injected current is added to the bias current supplied by V_B through R_L ; if the total current exceeds the peak current in Fig. 12, the diode switches to the high-voltage branch of the characteristic. The stable point on the low-voltage segment has been eliminated. The current through the diode is reduced by the switching (see Fig. 12), and some of the current is diverted to the output to serve as inputs to the following logic stages.

All of the disadvantageous aspects of logic with bistable devices are present. The lack of negation and the requirement of a separate reset operation are present. The outputs are not established by reference to V_B and ground potential, but depend on the characteristics of the circuit and on the value of the input signals. The sensi-

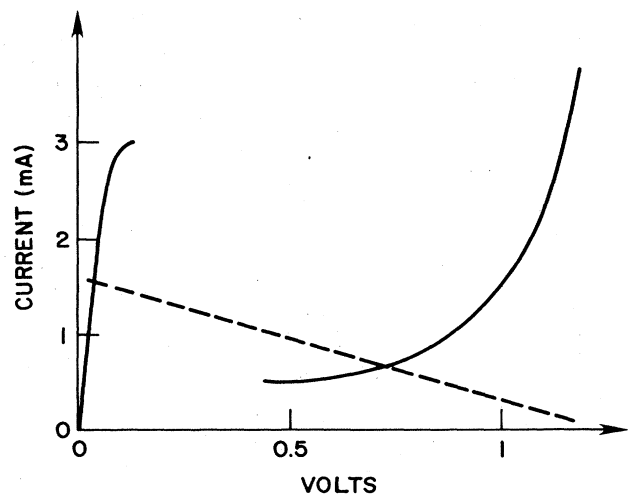


FIG. 12. Voltage-current relationship of the tunnel or Esaki diode. The dashed line is the load line determined by V_B and R_L in the circuit of Fig. 13.

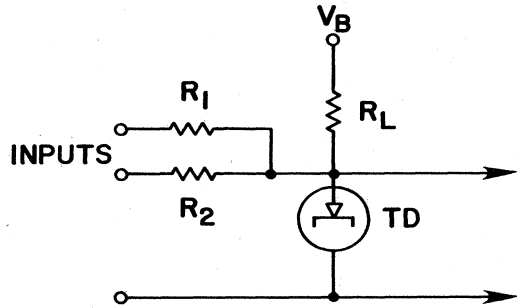


FIG. 13. Bistable circuit based on the tunnel diode.

tivity of the operation to variability in the tunnel diode characteristics and in the values of R_L and R_1 and R_2 demands high precision in these parameters. However, it proved difficult to contain values of the peak current within a range of a factor 1.5 in practice. The resistors R_1 and R_2 do not provide good isolation between stages. When the circuit of Fig. 13 switches, voltages change and signals propagate back through R_1 and R_2 to the preceding stage. The state of a logic circuit is influenced by the state of following stages, and not only, as intended, by its inputs.

In addition, the current output to following stages is derived from the bias and signal currents. Since there is a limited supply of this current, and since part of it passes through the tunnel diode rather than to the outputs, the ability of the circuit to provide fan-out to succeeding stages is limited severely. This aspect is com-

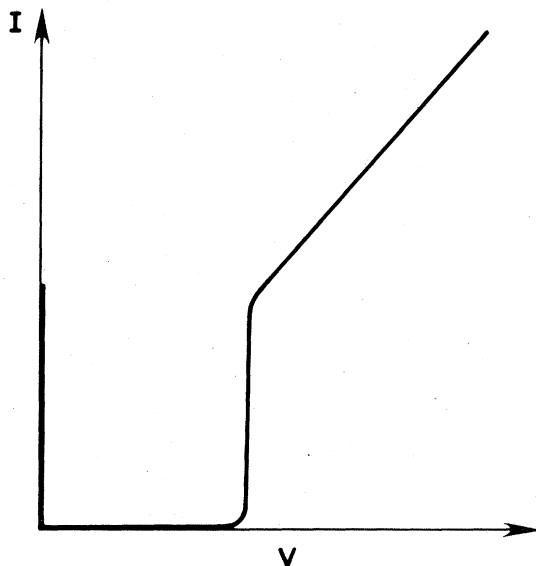


FIG. 14. Dependence of the voltage across a Josephson cryotron on the current that it carries. Note the zero-voltage current.

mon in bistable devices: the output energy is supplied by the bias and the input signals, and this input energy is attenuated in the device; so, little is available to provide fan-out. High current gain is lacking.

C. The Josephson cryotron

The current-voltage characteristic of a Josephson cryotron is shown in Fig. 14. A certain amount of current can flow through the junction with zero voltage across it. When the current exceeds the threshold, the cryotron switches to a voltage state. The resemblance to the tunnel diode, Fig. 12, is apparent, and a bistable circuit can be formed in a similar way, as shown by the dashed line in Fig. 14 (Matisoo, 1967). The circuit can be operated by current injection, in the same way as the tunnel-diode circuit. The same disadvantages accrue and need not be restated. The cryotron contains an additional feature, however. The maximum zero-voltage current, the threshold current in Fig. 14, can be decreased by application of a magnetic field to the circuit. Switching can be caused to occur by reducing the threshold to a value less than that of an existing current. The magnetic field can be produced by a current that is controlled by cryo-

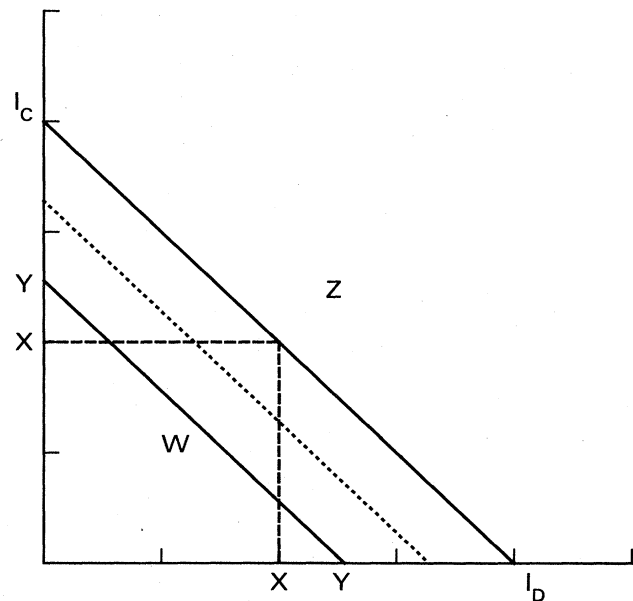


FIG. 15. Division of the state of a Josephson cryotron into voltage and zero-voltage states in the device-current-control-current plane. The AND function is performed by starting with a combination of currents that maintain the device in the zero-voltage state, the lower left portion of the plot. Increasing the currents in such a way as to move the operating point into the voltage state, above and to the right of the diagonal lines, for example, from W to Z , causes switching to the voltage state. The dotted line is the nominal threshold and the solid diagonals represent extreme values for its position. To guarantee performance of the two-input AND for all possible thresholds, the currents must lie between X and Y .

trons in a preceding logic stage. This additional possible way to cause switching allows great flexibility and ingenuity in the design of circuits. The circuit characteristics are often presented in another way, illustrated in Fig. 15. Here one axis represents the current through the device, and the other axis represents the current producing the magnetic field, designated the control current I_C . The dependence of the state of the cryotron on both of these currents can be described by a line that separates the plot into two regions—one containing points in the zero-voltage state, the other containing the voltage state. Switching is accomplished by any combination of changes in the currents that move the operating point across the boundary, say, from W to Z in Fig. 15. However, if the boundary between the two regimes is uncertain, as suggested in Fig. 15, the acceptable range of signal values can again be small. For example, if a two-input AND were implemented with Fig. 15, two currents greater than X will always cause switching to the voltage state; however, only one current greater than Y will also cause switching. The linear addition of currents causing a threshold to be exceeded again leads to the high sensitivity to parameter variations described in connection with Fig. 13 and the tunnel diode.

VII. CONCLUSIONS

The devices that have enabled large computing systems to be built have three terminals. This allows good isolation of the input from the output. Large nonlinear

changes in impedance in response to an input signal permit connection to standard values that represent binary digits. High nonlinearity provides large noise margins separated by a region of high gain and a tolerance of signal distortion and of the component variability that accompanies the large system environment and the limitations of low-cost production of large numbers of components. In contrast, the two-terminal bistable devices proposed as alternatives to the transistor demand tightly controlled component values that are incompatible with the economics of mass production and the variability of the environments in which they must operate. They also frequently show poor isolation of input from output, are unable to provide large fan-out, and require a separate resetting operation.

REFERENCES

- Gentile, S. P., 1962, *Basic Theory and Application of Tunnel Diodes* (Van Nostrand, Princeton), Chap. 8.III.
 Harvard University Computation Laboratory, 1946, *A Manual of Operation for the Automatic Sequence Controlled Calculator* (Harvard University, Cambridge, Massachusetts).
 Kohonen, T., 1972, *Digital Circuits and Devices* (Prentice-Hall, Englewood Cliffs).
 Lo, A. W., 1961, IRE Trans. Electron. Comput. EC-10, 416.
 Matisoo, J., 1967, Proc. IEEE 55, 172.
 von Neumann, J., 1958, *The Computer and the Brain* (Yale University Press, New Haven, Connecticut).
 Wallmark, J. T., and L. G. Carlstedt, 1974, *Field-Effect Transistors in Integrated Circuits* (Wiley, New York).

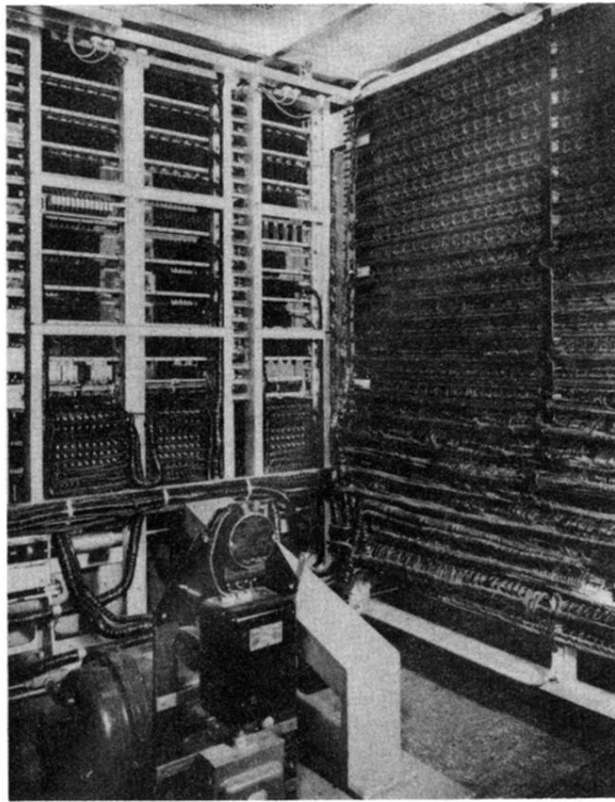


FIG. 1. View of the Automatic Sequence Controlled Calculator (from Harvard, 1946).

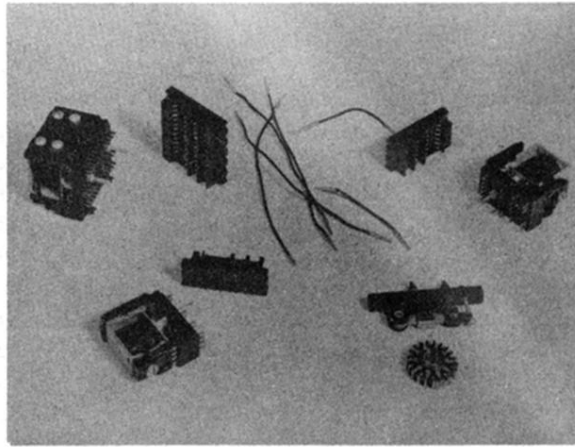


FIG. 2. Components used in the Automatic Sequence Controlled Calculator (from Harvard, 1946).