

THE FUTURE OF SOLID-STATE ELECTRONICS

**For three decades most measures
of the performance of digital electronics
have been improving exponentially.
What limits are in sight?**

Robert W. Keyes

The growth of commerce around the end of the 16th century created a need for large amounts of numerical calculation. The increasing understanding of astronomical phenomena around the same time also depended on accurate computation. The difficulty of accurately performing large amounts of tedious numerical work by hand stimulated Leibniz, Pascal and others to invent mechanical adding and multiplying machines in the 17th century. (See figure 1.) The speed and complexity of mechanical computation were limited by inertia, friction, size and the difficulty of transmitting information very far. The development of electrical technology led to the introduction of electrical components in peripheral parts of computing machines, such as motors in place of hand cranks. Major improvements in performance, however, had to await the electrical representation of information.

The earliest systems that could be called computers in a modern sense used relays to perform logic operations with electrical signals. These machines proved their worth by greatly speeding up the performance of a large variety of calculations and in turn stimulated a search for faster, more reliable components. The first commercial computers were built with vacuum tubes, which operated much faster than relays. The invention of the transistor and its substitution for the vacuum tube triggered the electronic revolution that is still under way. Although solid-state electronics has many facets and uses, its application to information processing has been the principal motivation for its long-continued and rapid development.

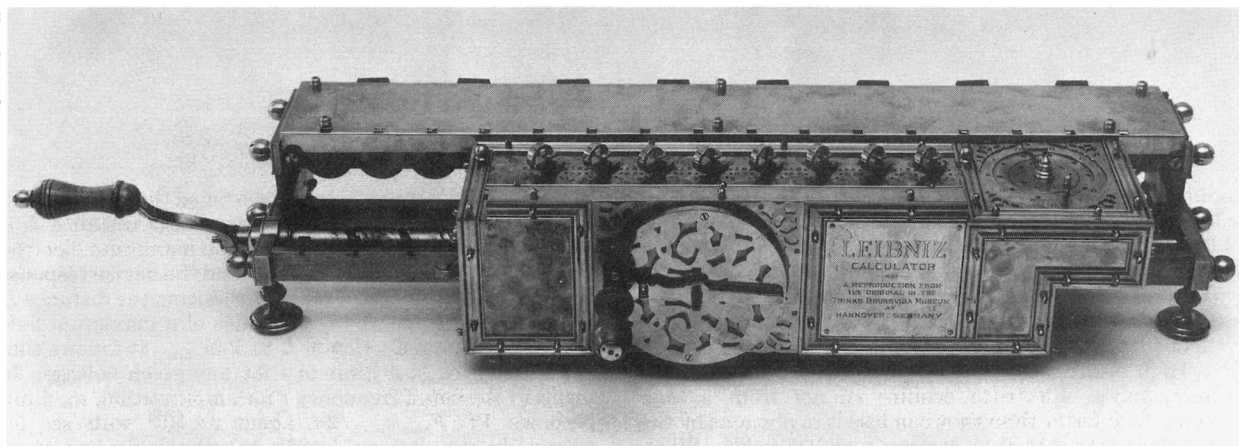
The quest for lower cost, higher speed, lower power consumption and higher reliability in computing equipment has been advanced through miniaturization and integration. Miniaturization, by decreasing capacitances and distances between components, leads to higher speeds and lower energy per operation. Miniaturization also has facilitated integration, the fabrication of large numbers of components on one chip. Integration is the key to low cost because the cost of producing a chip is only weakly dependent on its content. Also, the connections made by the methods of integrated electronics have proved to be much more reliable than discrete wires and solder joints.

Continual improvements in all measures of performance over the past 30 years have made it possible to construct fast, reliable machines containing tens of millions of components. One is compelled to wonder what lies ahead. How long can miniaturization continue? Are there more promising avenues for progress in the desired directions?

Digital computing and solid-state electronics

To answer such questions one must first know what a device in a computer must do and what is hard about its

Robert Keyes is a research staff member at the IBM Thomas J. Watson Research Center, in Yorktown Heights, New York.



The Leibniz calculator, invented in 1671. Information was transferred from one part of the machine to another by motion of the carriage. **Figure 1**

task. In automatic data processing, information is represented by the value of a physical variable. The information is handled many times in a computer, perhaps even hundreds or thousands of times. As a signal passes through devices and from one device to another, it has many opportunities to be degraded by attenuation and dispersion and to be contaminated with noise, cross talk and reflections. Deterioration of the information must be prevented by digital representation: The physical variable that represents a digit is reset to a standardized value at each step of a process so that a correct digital representation is always transmitted. Use of the binary system of numbers, with each digit a 0 or a 1, means that the recipient of a signal has to decide between only two alternatives. The use of electrical voltage has dominated as a means of representing information because the standard values for the digits are easily made available throughout a system and because signals can be sent long distances over wires to other parts of a system.

A logic operation is performed by controlling switches that connect an output to one of the distributed standard potentials. Figure 2 shows an example of switches in a computing circuit. If one or more of the switches is closed, the voltage seen at the output terminals *O* is zero. If all of the switches are open, then the output terminals are connected through the resistor *R* to the power supply and a voltage *V* appears at the output. If a closed switch and a positive voltage at the output are taken to represent 1's, and an open switch and zero voltage to represent 0's, then the circuit performs the logical function NOR. For purposes of computation the switches are controlled by electrical signals derived from preceding logic stages.

Clearly, the collection of tens of millions of switches into systems that can be made widely available requires that the cost of each switch be very low. The invention of the transistor inaugurated the era of solid-state electronics and provided the key to the eventual attainment of this very low cost, enabling the assembly of very large machines and the widespread use of computers that we see today. Solid-state devices proved to be extraordinarily suited to miniaturization and mass production through the well-known development of the integrated circuit. Most semiconductor fabrication today takes place on wafers that contain well over 10^9 transistors. The cost of processing a wafer can be divided among all of these transistors. The large numbers and the essential role of cost mean that electronic information-handling is a

technology rather than a science, and one may have to look beyond elementary physical principles to understand its limitations.

The present state of transistor technology is the result of three decades of research and development devoted to the processes and tools that make miniaturized transistors and high levels of integration possible. For the last two of these decades it has seemed incredible to many observers that the rate of progress, which could be gauged by plotting the logarithm of practically any measure of the technology against the year, could endure very far into the future. Figure 3 shows how the minimum size of features that can be manufactured on a chip has decreased by a factor of 2 every five years. There has been a continuing search for aspects of device physics or of fabrication processes that would pose insurmountable obstacles to further progress. Often when such limits have been suggested, however, the ingenuity of scientists and inventors has prevailed, and ways to circumvent the perceived limits were discovered. The search for limits, both physical and otherwise, continues to this day. (Some of the issues considered in this article are addressed in greater detail in reference 1.)

Limits set by electric fields

One source of limits can be found in the increasing electric fields in semiconductor devices as they are made smaller. Although miniaturization has been accompanied by a trend toward the use of lower voltages in circuits, the decreases in voltage have not kept pace with the decreases in linear dimensions. The reason lies in the existence of certain voltages, corresponding to energies that must be supplied to an electron, that characterize the response of semiconductor devices to applied potentials. One of these is the thermal voltage kT/e , which is 0.026 V at 300 K. For a barrier to act as a switch, its energy must be changed from a level that many electrons can cross to one that very few electrons can surmount. Because electron energies in semiconductors span a range of several kT , the barrier must be changed by much more than this energy uncertainty.

Another characteristic voltage derives from the energy gap of the semiconductor, manifested as the built-in voltage of p-n junctions. The built-in voltage times the electronic charge is the difference in band energies that arises from the different positions of the Fermi level with respect to the band edges in n and p semiconductors. The

built-in voltage constitutes a barrier to the passage of electrons across a junction; in silicon it is typically 0.5–1.0 V, depending on doping. A comparable voltage must be applied to cause a large current to flow across a junction. Similar considerations determine the gate voltage that is necessary to create a channel at a surface in a field-effect transistor.

High electric fields in semiconductors have several consequences. Electrons acquire energy from a high electric field faster than they can lose it to phonons by the scattering processes that dominate at low fields. When the average energy of these “hot” electrons becomes much larger than the kT of the lattice, new scattering mechanisms affect their motion and energy loss. The new scattering processes eventually lead to velocity saturation: The average electron velocity becomes independent of field, rather than proportional to field as it is with a mobility law.

At still higher fields the electrons attain energies that exceed the energy gap of the semiconductor, so that an electron can lose energy to an excitation across the gap, creating an electron–hole pair. The new carriers are also accelerated by the high field, and an uncontrolled avalanche of carriers and current results, possibly with destructive consequences. The condition for avalanche breakdown can be stated as a maximum allowable electric field in a reasonable approximation.

At an early stage of the era of solid-state electronics Edward O. Johnson recognized that these high-field phenomena lead to a basic limit on the performance of

semiconductor devices.² Johnson argued that the shortest time in which an electron can traverse a distance L is given by $t > L/v_{\max}$, where $v_{\max}$ is the maximum electron velocity. The time can be reduced and the device response extended to higher frequencies by reducing the distance L . For a given voltage V , the existence of a maximum field $F_{\max}$ limits the reduction of L to $V/F_{\max}$. It follows that $t/V > 1/F_{\max} v_{\max}$, a limit to t for any given voltage. In terms of the cutoff frequency f for amplification, the limit becomes $Vf < F_{\max} v_{\max}/2\pi$, about 2×10^{11} volts/sec for silicon. In almost three decades of technological progress, no one has found a way around this early limit.³

Another very durable limit was found some time ago by B. Hoeneisen and Carver A. Mead, who observed that miniaturization of semiconductor devices involves thinning the layers that separate regions of a device that are maintained at different potentials.⁴ The Fermi level is far from the band edges in these regions, and they are depleted of mobile charge carriers. Depleted layers occur, for example, in p–n junctions and between the substrate and the conductive channel in field-effect transistors.

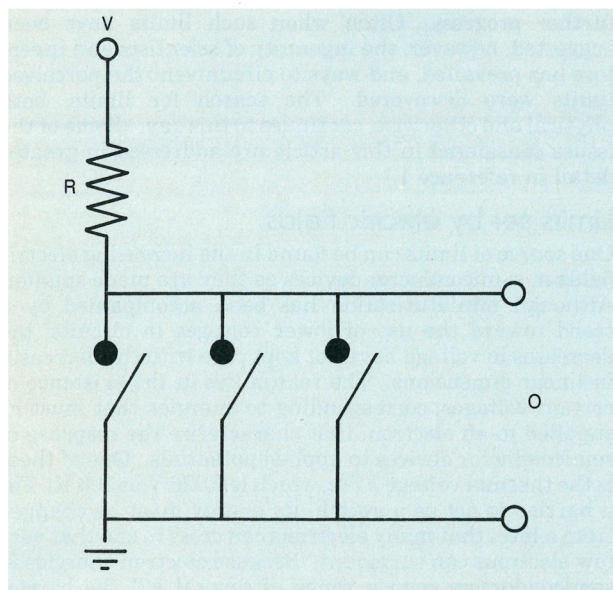
The variation of the potential in a depleted layer is controlled by Poisson’s equation applied to the charge density caused by the unneutralized donor or acceptor impurities. The voltage V supported by the layer, the thickness w of the layer and the impurity density N are related by $w^2 = 2\epsilon V/Ne$, where ϵ is the dielectric susceptibility of the semiconductor. Because of the difficulty of reducing voltage, the main burden of reducing w has fallen upon the impurity concentration N . Miniaturization of semiconductor components has been accompanied by a trend to heavier doping, and the electric fields in the depleted layers have increased.

The limit proposed by Hoeneisen and Mead is the point at which the layer can no longer withstand the voltage; currents flow across it either by tunneling through the layer or by accelerating electrons enough to stimulate an avalanche. The upper curve of figure 4 shows the limits on base thickness in a silicon transistor determined by Hoeneisen and Mead.

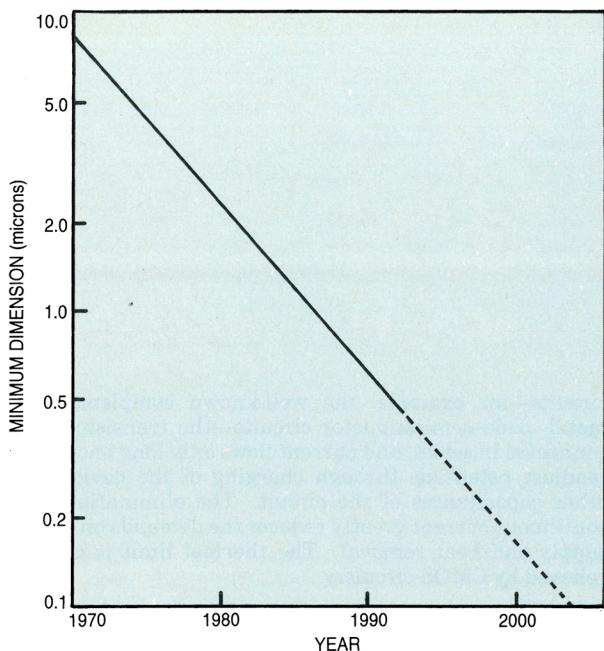
Another significant consequence of the heating of electrons in high fields is that it enables some electrons to surmount barriers that contain them at lower fields. This has become very important in the design of small insulated-gate field-effect transistors with SiO_2 as the insulator: A few electrons gain enough energy to enter the oxide, where they become trapped. Although the entry of an electron into the insulator is a rare event, the trapped charge accumulates with time and gradually changes the operating characteristics of the transistor, limiting its useful life in a circuit.

Other limits

Limiting effects are most readily analyzed for planar structures that are uniformly doped and can be treated as one-dimensional problems. But the processes used to fabricate small devices rarely produce constant doping throughout substantial regions, and the nonuniform doping and complex shapes of interfaces produce spatially varying electric fields and preclude the application of simple models. One can take advantage of these complexities in device design to deflect the limits estimated from simple models. Here the limited ingenuity of people in



Switches connected to perform the logic operation NOR with electrical signals. The switches are operated by electrical signals from preceding logic stages and connect the output O to potential V or to ground. All computers are based on such circuits, with transistors acting as switches. **Figure 2**



Miniaturization: Continuing development of photolithographic techniques has allowed the size of features that can be produced on the surface of a chip to be halved every five years. **Figure 3**

devising and making patterns of dopants acts as a limit.

Furthermore, new phenomena become apparent in the characteristics of very small devices. Part of the reason for this is illustrated in figure 5, which shows some length parameters that characterize silicon at 300 K. Dimensions of significant regions of devices can be less than some of these lengths, in which case novel effects can occur. This does not necessarily mean that the devices will not work, but that one must use different regimes of physical approximation in analyzing them. Electrons can pass through a small device in less time than it takes for their motion to achieve a steady state, resulting in a transient effect known as velocity overshoot.⁵ When the size of a region of a device becomes smaller than the mean free path of an electron, electrons can traverse the region without being scattered, an effect named ballistic transport.⁶ Although there is no doubt about the reality of these effects, they are difficult to quantify for actual devices, and limits derived for large devices cannot be straightforwardly extended to very small devices.

These limits are based on well-established solid-state physics, but they cannot be called fundamental because they depend on material parameters. Today they are mainly applied to silicon, and present knowledge suggests that other materials may offer substantially improved performance in terms of miniaturization and other functional parameters.⁷ For example, figure 4 shows that Hoeneisen and Mead's limit⁴ for the base width of a bipolar transistor is greatly reduced if silicon carbide is used in place of silicon. However, the existing investment in silicon processing methods and their advanced state of development precludes any early use of alternative materials in information processing hardware.

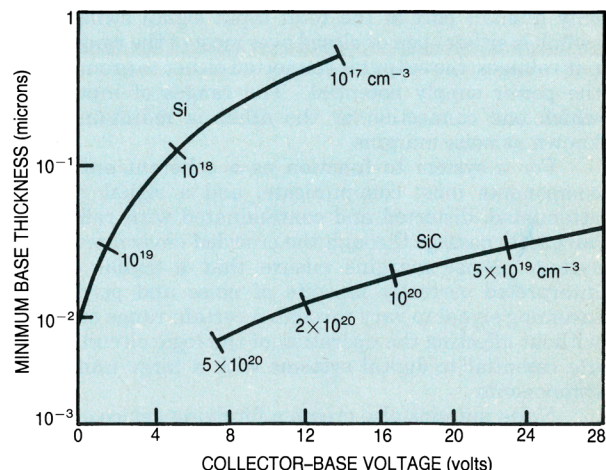
An increase in the density of electrical energy dissipation into heat is a further troublesome consequence of miniaturization. Capacitance is dimensionally a length: If all dimensions of a structure are reduced by a

factor z , then its capacitances are reduced by the same factor. Since the reduction of voltage in electronic circuitry is limited, the energy dissipated in charging and discharging a capacitance during circuit operations also decreases as z . The density of components on a chip, however, increases as $1/z^2$, so the density of heat production increases as $1/z$. The situation can be even worse than this because ingenious component designers can increase circuit density even faster than scaling alone does. Also, miniaturization is often accompanied by an increase in circuit speed, so that capacitances are charged and discharged more times per second as dimensions are reduced. However, there is a compensating factor: Tricks in the design of components and changes in materials can decrease capacitance without a concomitant scaling of dimensions.

In spite of these perturbations of a simple trend, cooling technology has been challenged by a steadily increasing production of heat per unit area, especially in machines of the highest performance. There is undoubtedly some kind of a maximum cooling rate, but up to now cooling technology has kept up. The actual numbers in large computers are startling: 30 watts may be produced in a chip with an area of 1 cm^2 . For example, 30 W/cm^2 is the power radiated by a blackbody at 1500 K, and surfaces used for cooking have only a tenth of this power density. Nevertheless a limit due to cooling requirements is not imminent: The removal of 1000 W/cm^2 from a silicon chip has been demonstrated.⁸

The transistor as a switch

The perception of limits on the currently dominant silicon devices encourages the invention of alternative technologies. Physicists are prominent in this endeavor. (The



Limitations on base width in a bipolar transistor arising from junction breakdown. The indicated doping concentrations are needed to avert punch-through—the total depletion of the base. Its material properties give silicon carbide a much greater potential for miniaturization than silicon has. (Adapted from ref. 1.) **Figure 4**

intimate connection between physics and electronic devices can be seen in the October 1986 and February 1990 special issues of *PHYSICS TODAY*.) Novel solid-state phenomena readily tempt their discoverers to envisage new device concepts that will soon relegate transistors to obsolescence. Before we look at such possibilities, however, let us try to understand what the transistor does and why it has been so successful.

The physics of transistor action is amply described in many books.⁹ In addition to its susceptibility to miniaturization, the transistor has excellent properties as a switch. Transistors can be made to have very high gain; that is, the signal that controls the switch can be much smaller than the current and voltage that are switched. This high gain is the most important point to be made about the transistor concept: Because a few immobile electrons can neutralize mobile holes that can carry current for as long as desired, a charge placed on an electrode is able to control 100 or 1000 times more charge. The same principle—placing charge on an electrode to allow other charge to flow by—was also the basis of the triode vacuum tube, and so has been the basis of electronic amplification for the better part of a century.

When used as a switch in a logic circuit, a transistor has either a high conductance (switch closed) or a very low conductance (switch open). High gain means that the transition from one output state to the other occurs over only a small part of the total input signal swing: The switch is either open or closed over most of the range of input voltages; the output is connected either to ground or to the power supply potential. The ranges of input over which one connection or the other is maintained are known as noise margins.

For a system to function as a coherent entity the components must communicate, and a signal may be attenuated, distorted and contaminated with reflections during its passage through the crowded environment of a system. Noise margins ensure that a signal can be interpreted correctly in spite of noise and permit the incoming signal to vary through a certain range of values without affecting the operation of the logic circuit. They are essential to digital systems with a large number of components.

Noise margins also create a forgiving device environment: The exact value of the gain and the range of inputs at which it appears are not critical parameters. Such tolerance of device variability is necessary for the success of a low-cost, highly integrated technology.

The availability of transistors of two polarities is another great advantage of transistor circuitry. The word "electron" could be replaced with "hole," and vice versa, in all of the preceding text. The existence of two polarities has made possible the "complementary" concept, in which the same signal turns on one transistor but turns off another transistor of opposite polarity. In complementary

circuits—for example, the well-known complementary metal oxide–semiconductor circuits—the transistors are connected in series, and current flows only long enough to readjust potentials through charging of the device and other capacitances of the circuit. The elimination of a continuous current greatly reduces the demands on power supply and heat removal. The thermal limit is greatly relieved by CMOS circuitry.

Negative resistance and other devices

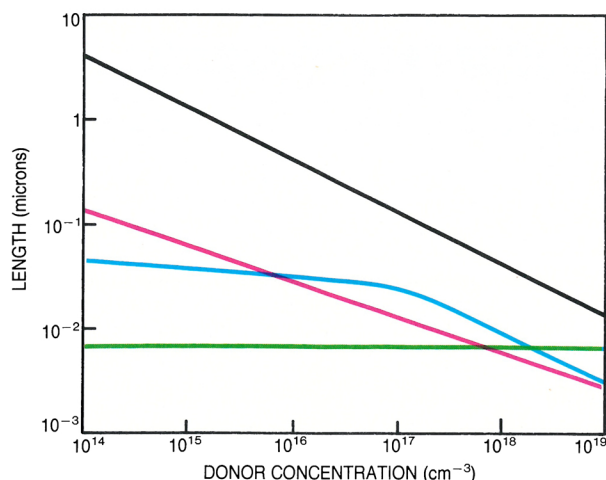
In the search for alternatives to transistors, imagination readily stretches an irregularity in a current–voltage characteristic into a new way of doing logic. In particular, one can use a negative resistance to construct a bistable circuit, and the means to switch the circuit from one state to the other are easily provided. Indeed this idea was once the focus of a considerable body of research and development activity; bistable circuits based on the negative resistance of tunnel diodes were widely believed to offer an opportunity to construct very fast logic circuitry.¹⁰

The basic idea is illustrated in figure 6, which shows how the stable states (labeled A and B) of a negative-resistance device in series with a resistor are determined. If the circuit is set initially to state A, and a small additional current is injected to make the total exceed the low-voltage peak of the characteristic, then the circuit will switch to the other stable state, B.

A major problem with this scheme is that high gain—that is, switching with a small input current—depends on state A being very close to the peak. That is, the gain is limited by the ability of technology to control the location of A and the peak. While one can easily adjust these parameters to demonstrate operation of the circuit in a laboratory, matching them closely enough to achieve useful gain for all of 10^5 or 10^6 devices that have been mass-produced to make a large system is another matter. Furthermore, the parameters of solid-state devices depend on their temperature, and their gains will vary with the different environments in which the systems are used.

Another difficulty of the bistable circuit is that its action is not that of a switch: The output signal is determined not by connection to a standard but by the device and will vary with the device characteristic.

Logic with optical bistability has also received attention recently but is plagued by all of the same problems: the intimate dependence on device parameters, the absence of comfortable noise margins because of the need for critical adjustments to obtain gain, and the lack of a standard that can be used to determine output values. In fact, the last disadvantage is not associated with a particular optical device concept. There does not seem to be any practical method for establishing, distributing and making connections to an optical standard that is comparable to the distribution of voltages around a system by some form of metallic wire.



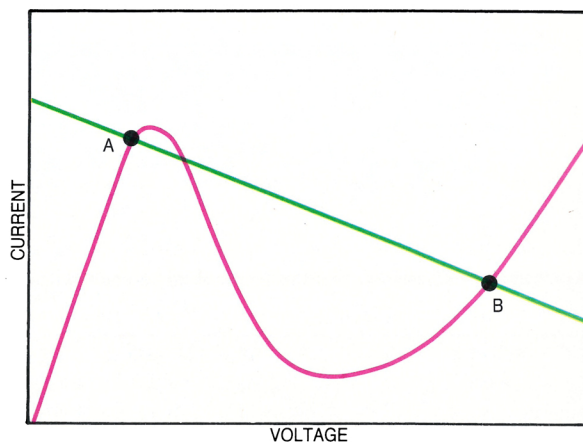
Characteristic lengths in n-silicon at 300 K as a function of donor concentration. Shown are the minimum depletion-layer thickness that will support 1 V (black), the mean free path between scattering events (blue), the average distance between donor atoms (red) and the average electron wavelength (green). (Adapted from ref. 1.) **Figure 5**

Another interesting idea, stimulated by recent advances in mesoscopic physics, is based on interference of electron waves in solids.¹¹ Such devices depend on the ability to fabricate electronic structures with conductors so small that electron waves are coherent throughout them. In one embodiment of a device an electron wave is divided between two identical parallel channels. After traveling a certain distance the two waves are allowed to recombine and interfere to reconstitute the original wave. A proper signal applied to a third electrode can change the velocity of propagation in one channel so that the waves arrive at the output terminal out of phase and interfere destructively, producing no output.

Again, this kind of device does not offer noise margins. The signal applied to the third electrode must be just right; if it is too large or too small the cancellation will be incomplete. And the output is sensitive to the dimensions of the device, because the desired cancellation at the output will occur only if the lengths of the paths are correct. There is little room for error in fabrication. The device can be demonstrated in a laboratory, where the necessary adjustments to the electrical inputs can be made, but is not compatible with the requirements of technology for large systems.

The future of semiconductor technology

The steady introduction of new tools and techniques into the processing of semiconductor wafers has sustained the long-continued miniaturization of semiconductor components. This has necessarily been accompanied by refinement and evolution of all aspects of device technology. There is no doubt about the physical reality of the effects I have cited as limits to the continuation of these trends. The limits derive, however, from simple models: large devices, planar structures divided into homogeneously doped regions, and potentials acting long enough to achieve a steady state. Therefore the limits are fragile; as devices are reduced in size they enter new regimes of



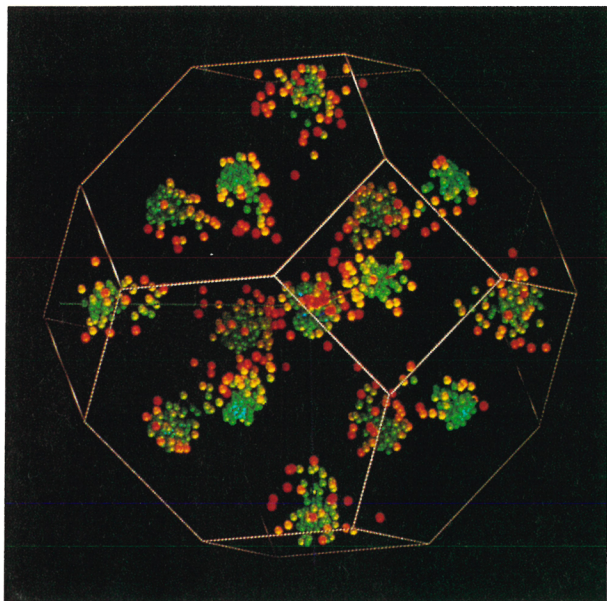
Negative-electrical-resistance device (such as a tunnel diode) with the characteristics shown could be used to make a bistable circuit for doing logic. The diode (red characteristic) is connected in series with a resistor that determines the load line (green). The circuit is initially biased at point A. Switching is accomplished by injecting additional current from a preceding stage that raises the current through the diode above its low-voltage peak, causing the operating point to move to B. **Figure 6**

device physics. Can we peer into the future and anticipate how the difficulties and apparent limits to miniaturization will be confronted?

More or less traditional lines of development—devising new silicon structures that will function at smaller dimensions and reduced voltages—will certainly continue. But it is increasingly difficult to understand what actually happens in these small structures with complex inhomogeneities. Tractable analytical models cannot take into account the large two- or three-dimensional variations in material parameters that may be encountered within a mean free path, the strong dependence of scattering on electron energy, the finite times required for steady-state conditions to be established in a device, and nonparabolic band features (that is, deviations from $E \propto p^2$) at the energies attained by carriers in miniature devices.

Numerical methods, made increasingly practical by the rapid increases in available computing power, are valuable in replacing analytical models and supplementing intuition and empiricism in the development of optimal device designs. Monte Carlo methods, for example, follow individual electrons through a device and can take into account complex dependences of scattering probabilities and other features.^{12,13} The ability of simulation to provide insight into the new physical contexts and to aid in the formulation of ideas and points of view may be even more productive than its direct application to device design. For example, figure 7 shows the calculated excitation of electrons into remote parts of the Brillouin zone, which one must consider when thinking about small devices.

Like other facets of electronics, simulation has limits. It is, of course, always limited by the computational power available. The more electrons that can be followed, the more accurate the result. Including inhomogeneity in more detail would bring the simulated structure closer to the real one. A very large expansion of computational



Hot electrons invade the high-energy portions of the Brillouin zone of a zinc-blende-type semiconductor in a simulation of a GaAs metal-semiconductor field-effect transistor (MESFET). In the absence of an electric field the electrons occupy an energy minimum at the center of the zone. Different colors indicate different kinetic energies with respect to local energy minima. (Copyright 1990 by International Business Machines Corp; reprinted with permission from ref. 13.) **Figure 7**

power would permit full three-dimensional Monte Carlo experiments. The accuracy of simulation, however, is also limited by our incomplete knowledge of all relevant aspects of semiconductor physics. The dynamics of electrons and holes throughout the range of energies that are found in miniaturized semiconductor devices has never been thoroughly explored. It is not an easy subject; important events take place on subpicosecond time scales and are difficult to probe.

The fabrication of small complex structures is also a source of limits, and is not helped by the inability of analytical tools to probe composition with sufficient spatial resolution. The power of computers is also used to attack this problem, through simulation of steps in fabrication processes.¹⁴ Process models attempt to follow impurities from their entry into a material by diffusion, ion implantation or some deposition process, through treatments at elevated temperatures, and also seek to take into account the changes in surfaces that result from interactions with their environment. Doing all of this in three dimensions is desired. There are many phenomena that must be considered: interactions among impurity atoms and between impurity atoms and crystal defects, the effect of dopants on the Fermi level, the effect of the Fermi energy as a driving force for atomic motion, and precipitates and reactions at surfaces that can act as sources and sinks for point defects and dopants.

To an even greater extent than the modeling of device performance, the modeling of device fabrication is limited by incomplete knowledge of the underlying basic physics—the mechanisms of diffusion, interaction and reac-

tion. Even in silicon, an exceptionally well-known material, the roles of interstitial atoms and vacancies in diffusive processes are still disputed. Enhancement of diffusion in the vicinity of a surface that is reacting with oxygen is well known and important in silicon processing, but the underlying detailed atomic mechanisms have not been established. The same may be said about other sources and sinks for defects. The recent incorporation of germanium-silicon alloys into silicon device structures further complicates the picture.

The last remark introduces a revolution that is well under way: the invasion of device technology by epitaxy and heterostructures. For a long time we have considered a device to be something fabricated in a single semiconductor material. The advent of heteroepitaxy means that different semiconductors with different energy gaps and dielectric constants may be encountered within distances of a few nanometers. The use of alloys even allows gradients of energy gaps in addition to variable impurity concentrations to be built into devices. The variation of lattice size among different semiconductors offers strain produced in heteroepitaxy as another parameter that can be used in device design. Numerical modeling and computational resources have not yet attacked the inclusion of these additional variables—in part because the accumulation of the basic knowledge that should underlie simulation is even further away for alloys than for simpler materials.

The continual infusion of new knowledge and new methods into solid-state electronics has sustained the drive toward more powerful, more available information handling products for three decades. The tools to produce the hardware of the next decade of electronic technology are already in the development laboratories. A heavy dependence on modeling and simulation is essential to optimum use of the new technological capabilities. The finite ability of condensed matter physics to provide the basic knowledge on which the simulations depend is perhaps the “fundamental” limit to the advance of electronics.

References

1. R. W. Keyes, *Contemp. Phys.* **32**, 403 (1991).
2. E. O. Johnson, *RCA Rev.* **26**, 163 (1965).
3. M. Nagata, *IEEE J. Solid-State Circuits* **27**, 465 (1992).
4. B. Hoeneisen, C. A. Mead, *Solid-State Electron.* **15**, 819, 981 (1972).
5. G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, E. Ganin, S. Rishton, D. S. Zicherman, H. Schmid, M. R. Polcari, H. Y. Ng, P. J. Restle, T. H. P. Chang, R. H. Dennard, *IEEE Electron Device Lett.* **8**, 463 (1987).
6. M. Heiblum, L. F. Eastman, *Sci. Am.*, February 1987, p. 64.
7. R. W. Keyes, in *Silicon Carbide 1973*, R. C. Marshall, J. W. Faust Jr, C. E. Ryan, eds., U. South Carolina P., Columbia, S. C. (1974), p. 534.
8. D. B. Tuckerman, R. F. W. Pease, *IEEE Electron Device Lett.* **2**, 126 (1981).
9. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., Wiley, New York (1982). J. T. Wallmark, L. G. Carlstedt, *Field-Effect Transistors in Integrated Circuits*, Wiley, New York (1974).
10. R. H. Bergman, M. Cooperman, H. Ur, *RCA Rev.* **23**, 152 (1962).
11. S. Datta, M. R. Melloch, S. Bandyopadhyay, M. S. Lundstrom, *Appl. Phys. Lett.* **48**, 487 (1986).
12. C. Jacoboni, L. Reggiani, *Rev. Mod. Phys.* **55**, 645 (1983).
13. S. E. Laux, M. V. Fischetti, D. J. Frank, *IBM J. Res. Dev.* **34**, 466 (1990).
14. P. M. Fahey, P. B. Griffin, J. D. Plummer, *Rev. Mod. Phys.* **61**, 289 (1989).