31 January 1969, Volume 163, Number 3866 SCIENCE

Large-Scale Integration and the Revolution in Electronics

The rapid evolution of solid state electronics is having a major impact on the computer.

S. Triebwasser

The two decades since the invention of the transistor, in 1948, have seen a rapidity of evolution in electronics that can be termed revolutionary. The most dramatic impact of this evolution has been on the computer industry, which could not have grown so rapidly had not the new solid state technology been available. This growth, in turn, made possible the broad advances in space exploration and air travel, by providing both the tools for the computer analyses that were required during the design and construction phases and the ground and on-board electronics required in flight. The ubiquitous transistorized radio is clear evidence that the revolution is not solely in advanced technological areas; there has been a revolution in the commercial market as well, typified by an almost one-for-one replacement of the vacuum tube circuit by the transistor circuit with equivalent function.

Large-scale integration (LSI), the emerging form of solid state digital electronics, has particular application to computer-function requirements. Large-scale integration is a term used to describe the technology which consists of arrays of logic or memory, cells formed in a batch process to realize a complete function. The number of circuits involved in such an array may be 50 to 100 or more. At present it is only in computers or in systems performing computer-like functions (as, for example, in large-scale switching networks) that development cost for this kind of electronics is justified. Hence, large-scale integration and the revolution in electronics are discussed here in the context of computer technology. The new technology promises a much broader application of electronics to our everyday lives, but, as might be expected, it has generated new problems. Here I review these promises and problems.

Advances in the computer industry have been paced by developments in the new electronics that have occurred since the invention of the transistor. World War II provided a great stimulus to the widespread application of electronics in many areas. An enormous effort went into the miniaturizing of electronics equipment for military applications. Miniaturization was dramatically simplified by the availability of solid state components, which made possible revolutionary improvements in the reliability, size, and power consumption of electronic equipment. A further dramatic change has been a reduction in cost well below what would have been possible with vacuum-tube technology. These changes have had a major impact on the computer industry—an industry which, on the one hand, has stimulated the rapid development of transistors, integrated circuits, and, finally, large-scale integration and, in turn, has been made possible by these developments. The television industry, on the other hand, is still, for the most part, using vacuum-tube technology, since vacuum-tube sets remain competitive, in cost and reliability, with equivalent transistorized sets.

A new generation of computer equipment has appeared approximately every 5 years (*1*). With each succeeding generation there has been a reduction by a factor of 10 in the cost per arithmetic operation (*2*) and, at the same time, a reduction, by a factor of about 10, in the rate of failure of components, as measured by the usual standards. The reduction in the cost per operation has rapidly widened the range of problems that can be handled economically by electronic data-processing equipment, hence has rapidly widened the market, while the increase in reliability has made it possible to build larger systems without the pain of constant component failure. The extremely rapid buildup of demand for computational power and the fact that computers can be built from relatively few and conceptually simple digital logic circuits used in enormous repetition has made the technological development rapid as well. The impact of the development of solid state components is clearly evident from the fact that the second generation of computer equipment, which appeared at the end of the 1950's, was already completely transistorized. Advances in space and aircraft electronics have paced the rate of development of the space and aircraft industries, where the smaller size, lower power requirement, and increased reliability of solid state components are of obvious importance.

Let me briefly review these generations of computers. In about 1954, first-generation commercial computers, based on vacuum-tube technology, were being delivered to customers. A typical system had 2000 logic circuits in its central processing unit, with a mean time to failure, per circuit, of 1 percent per 1000 hours. The equipment was bulky, used much power, and failed much too often. The second-generation computers appeared at the end of that decade; these were transistorized, but individual transistors and resistors were

The author is manager of solid state electronics at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.

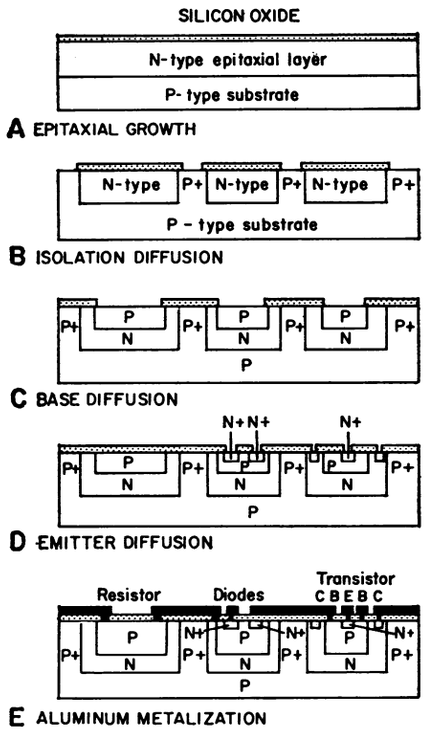


Fig. 1. Steps in integrated circuit processing.

wired in much the same manner as tubes. However, power consumption was considerably reduced, and rates of failure were down by a factor of 10, so it was possible to use many more circuits without prohibitive "downtime." Cost was reduced by perhaps a factor of 3, and performance was improved by a factor of 3 or better, depending on the size of the computer. By 1964, third-generation computers were being made with integrated circuits, using largely "hybrid" integrated circuit technology, although there was already some application of "monolithic" integrated circuits. The distinction between the "hybrid" and "monolithic" forms of integration is elaborated below. The primary point is that in either case (hybrid or monolithic) the building block was still the single logic circuit. Computers were being designed much as they were in the vacuum-tube days because the individual logic circuit was still the basic building block available to the designer. As an example of hybrid integrated circuit technology, in the currently employed IBM Solid Logic Technology (*3*), a logic circuit is mounted on an alundum (*4*) substrate about ½ inch (1.25 centimeters) on a side, with several transistors attached, the latter being hardly visible as compared with the ceramic carrier. By 1965 and 1966, while the single circuit was still the basic element in equipment being delivered for commercial use, in various laboratories the extension of monolithic integrated circuits to large-scale integration in the form of 100 or so interconnected circuits on a piece of silicon 0.1 inch on a side was being accomplished, as compared with one circuit on a carrier ½ inch square in the hybrid technology. The technology that is making this dramatic change possible will be discussed in the next section.

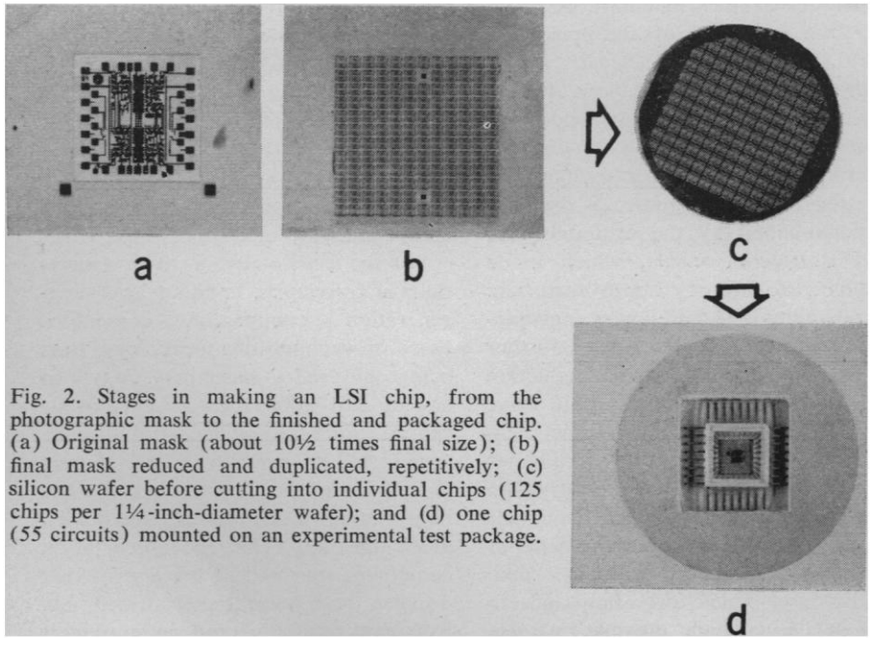


Fig. 2. Stages in making an LSI chip, from the photographic mask to the finished and packaged chip. (a) Original mask (about 10½ times final size); (b) final mask reduced and duplicated, repetitively; (c) silicon wafer before cutting into individual chips (125 chips per 1¼-inch-diameter wafer); and (d) one chip (55 circuits) mounted on an experimental test package.

Silicon Planar Technology

The current revolution in electronics is being made possible by the wedding of processes, such as diffusion and chemical etching, introduced early in transistor technology and photolithography, which had been developed for entirely different purposes. This marriage has led to the silicon planar technology (*5*). Some basic description of this technology is necessary for an understanding of what has been, and will be, happening. First, in order to produce the transistors and diodes that are required to make electronic circuitry, it is necessary to produce closely spaced regions in semiconductor material, with electrons as current carriers in one region and holes (absence of electrons) as carriers in the adjoining regions. It would be inappropriate, here, to elaborate much beyond this point; suffice it to say that this geometric arrangement of closely spaced regions is required. The electron-or-hole conductance behavior is produced by the introduction of impurities (negative, or N-type, and positive, or P-type) into the host semiconductor. Typically these N-type and P-type impurities could be phosphorus and boron, respectively, and the host semiconductor in current technology is silicon. The required impurity concentrations for various regions range from less than 1 part per million to 1 part per hundred. These impurities are introduced by means of diffusion, at high temperature, from the gas phase or by a technique known as epitaxial depositions of the "doped" material. Processes are applied to a piece of silicon in the form of a wafer of the order of 0.01 inch thick and from 1 to 3 inches in diameter.

So far, I have been talking about the kinds of processes that were in use before the photolithographic techniques were introduced. By "photolithographic material" is meant material which can be polymerized or depolymerized by radiation. In its polymerized state the material is much more resistant to chemical etching than it is when nonpolymerized. The result is that it is possible, by ordinary photographic exposure of complex patterns, to produce regions which are essentially impervious to chemical attack and others which can be etched easily. The final requirement is an insulating material which adheres well to the semiconductor and to which both photolithographic materials and metal films adhere.

The technology is shown diagramatically in Fig. 1. Figure 1A shows a P-type wafer on which N-type impurity material from the vapor phase has been deposited epitaxially. The term *epitaxial* refers to the fact that both regions are single-crystal and that the deposited region continues the lattice determined by the original wafer. The next step required to form the remainder of the structure of Fig. 1A is heating of the wafer in an oxidizing atmosphere; this causes SiO_2, a glassy impervious insulator, to grow. A layer of photoresist material is then deposited on the structure of Fig. 1A and exposed to ultraviolet light in the regions shown dotted in Fig. 1B. A two-step etching process follows. The first etch removes the nonexposed areas of photoresist; the second, and more powerful, etch removes the SiO_2 but not the polymerized photoresist, and thus cuts holes in the SiO_2 overlayer, as shown in Fig. 1B. Now the polymerized photoresist is removed, in preparation for the high-temperature step, in which the wafer is placed in a furnace at a temperature of about 1000°C in the presence of a boron-bearing atmosphere. After some time, enough boron atoms diffuse into the surface of the exposed silicon and down into the silicon to give, eventually, the structure shown in Fig. 1B. The SiO_2 is relatively impervious to the diffusant and protects the silicon surface, so that none of the boron diffuses into the region where the SiO_2 has remained. The doped regions extend some distance under the SiO_2 overlayers because the diffusion process proceeds both down into the material and sidewise. This sidewise diffusion must be taken into account in the design of the processes required to make the final structure. The reader can imagine the repeated photolithographic steps and diffusions that are required in order to make the structure shown in Fig. 1, C and D. The notation P+ and N+ designates regions of heavy doping.

The final photolithographic step is performed on a metallic layer which has been evaporated over the entire wafer. This step is preceded by a photolithographic step in which the SiO_2 is removed in regions where metallic contact is to be made to (Fig. 1E) the emitter (E), base (B), and collector (C) of a transistor and to the ends of a resistor. In the final step, the wafer is placed in a vacuum evaporator and entirely covered with a thin film of aluminum. Again, photoresist is applied, and photographic exposure is followed by etching to remove the metal in areas where contact or conductivity is not desired. We can see that this step satisfies two requirements. It provides contacts to the resistors, diodes, and transistors, and, in addition, this metal layer can be used to make interconnections between the elements. The interconnecting lines run over the circuit elements on the surface of the insulating layer of SiO_2 that overlies most of the silicon wafer. The availability of all contacts on the top surface of the wafer (from which the designation *planar* derives) represents a significant breakthrough. Herein lies the basis of the rapid development of integrated circuits over the past 5 years. We have progressed from (i) using the final photolithographic step to form contacts to individual regions of the transistor, through (ii) using this step to form interconnections among devices, to (iii) using it to provide the interconnections among circuits as well, to make large-scale integrated arrays. Where do we stop? Obviously, we could put down a complex pattern that would interconnect all circuit elements of the silicon wafer, which may number as many as a million.

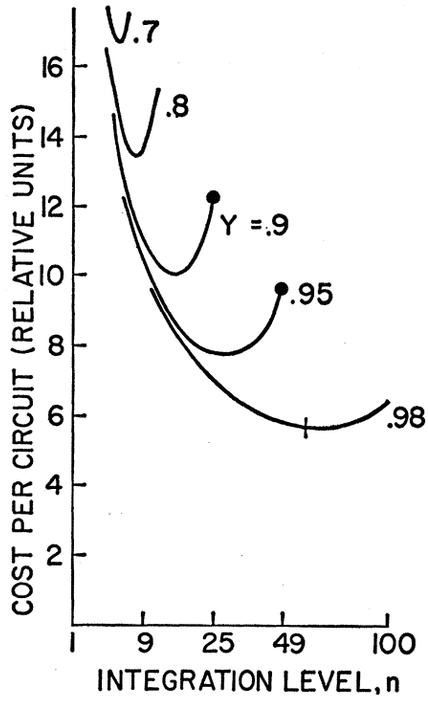


Fig. 3. Relative cost per circuit, plotted as a function of level of integration.

The extent to which we integrate is limited by two kinds of constraints. The first is technological and is based on the fact that none of the processes we use can be carried out with 100-percent yield. The second is more subtle and is related to the economics of the application of large-scale integration to the electronics market in competition with other forms of hardware.

Technological Limits

Figure 2 shows the basic steps required to make an LSI "chip." The question we examine in this section is, "How much can one chip contain?" As indicated above, this is determined primarily by technological considerations. These have to do with the quality of our materials and the skill with which we use them. The chip shown in Fig. 2 has a final size of less than 0.1 inch on a side. The dimensional control toward which we must work is measurable in microns. Photoresist layers and insulating layers and metalized layers which pile on one another are also of the order of 1 micron thick. Chemical etching procedures, then, must be very carefully controlled to cut down through these layers without substantial loss of definition. In addition, we are reaching the theoretical limit of optical exposure equipment, because of the fact that the dimensional control required is of the same order as the wavelength of light. These considerations lead to the results shown in Fig. 3.

If we regard the single circuit as the basic building block (a circuit consisting of between five and ten components, occupying on the order of 20 square mils on the surface of the silicon wafer), then Fig. 3 shows the manufacturing cost per circuit as a function of the number (n) of circuits per chip, with the circuit yield as a parameter. The cost per circuit goes down initially as the level of integration increases, because so much of the expense previously required to interconnect discrete components and circuits is now absorbed in a single evaporation step (compare this with having a box full of transistors and resistors and people with soldering irons connecting them to make a 20,000-circuit central processing unit). The cost goes through a minimum and turns back up because of the defects of our photographic and semiconductor processing technology, which lead to failures or rejected circuits. On the basis of a simplified model (*6*), we can say that, if the single circuit yield is Y, then the probability that the chip containing n

circuits is "good" is Y^n, if fault clustering is ignored. If we imagine Y to be ½, then it is clear that an attempt at making n of the order of 10 would produce very few chips that pass inspection. Historically (over the past several years), the minimum point on the curve shown in Fig. 3 has been moving down and to the right at a much more rapid rate than most of us were anticipating at the beginning of this decade. Projections at that time indicated that Y would by now be of the order of ½, but in fact, on the basis of more sophisticated models (and including fault clustering), it appears to be running above 0.95 for most of the kinds of circuits we are making today.

In fact, several companies have made, or are now making, experiments directed at circumventing the "Y to the nth" problem. Through preliminary testing of the single circuit and the use of computer programs to determine the assignment of the good circuits to particular places in a logic list and then to calculate wire-routing paths among those good circuits we could bypass the circuits which fail inspection (*7*). In a period when yields were more limited, this approach looked like good strategy for making large-scale integration systems, but now that single-circuit yields are very high, the additional inspection and special problems involved in this discretionary or programmed wiring approach make it less advantageous. It should be pointed out that a finished chip must be all good, otherwise it must be discarded, because there have been no economically practical means proposed for repairing defects, even if the defects could be easily identified.

Figure 4 illustrates the kinds of analyses that are being made in semiconductor laboratories. The two curves refer, respectively, to the bipolar transistor, which is the one in common use, and the insulated gate field effect transistor (IGFET), which is a conceptually and physically simpler device requiring about half the number of photolithographic steps required by the bipolar transistor. This greater simplicity is reflected in the fact that, with a given density of defects in photoprocessing, it is possible to integrate at a higher level with the IGFET than with the bipolar device. Laboratories and manufacturers of high-quality devices are operating at the 200-defects-per-square-inch level, but anticipate rapidly attaining the 100-defect level. The value of 20 percent for the percentage of chips

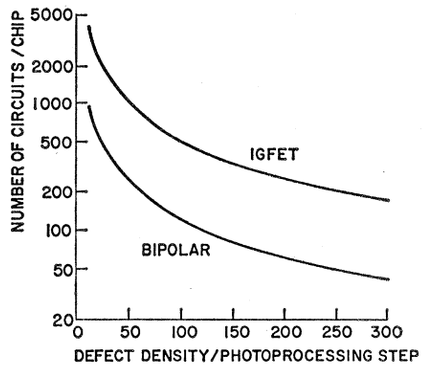


Fig. 4. Integration level achievable for a 20-percent chip yield (that is, for a yield in which 20 percent of the chips are good) at various defect densities (in numbers of defects per square inch).

that will be good was chosen because analysis of the trade-off of the chip cost, handling cost, and packaging cost indicates that this is an advantageous integration level at which to operate (*8*).

Figure 5 shows the level of integration as a function of time for the period 1960–70. The flat portion of the curve represents the era of single components (or the era of less than one circuit per chip); there is a rapid buildup after 1964. The levels of integration for the curve labeled "high performance" reflect the fact that specifications in the high-performance area are such that, at any point in time, Y is lower than it is for less stringently specified circuitry, and that the lower Y is, in turn, reflected in the level of integration attainable.

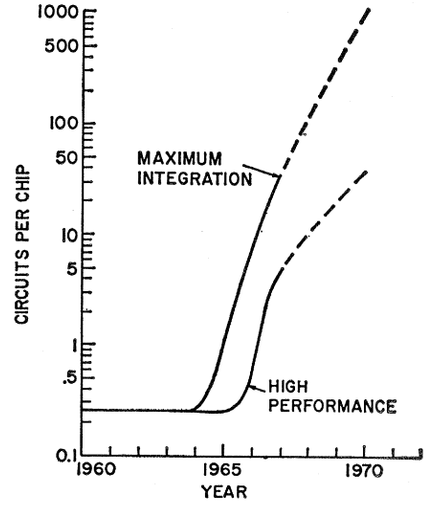


Fig. 5. Level of integration plotted as a function of time for the period 1960–70. [From W. A. Notz, E. Schischa, J. L. Smith, M. G. Smith, *Electronics* **40**, 130 (1967), with permission]

Systems Considerations

We will achieve lower-cost hardware, improved performance, and increased reliability only if we learn to take advantage of large-scale integration in implementing systems. Here we have to address the problem of what we should put on a chip. Figure 6 shows two masks of a set of four required to make a chip in field-effect transistor technology, comprising 122 logic circuits interconnected to perform part of the control function of the small computing system. In final size, these are less than 0.1 inch on a side. The problems associated with the design, verification, and testing of such a complex array of logic circuits are formidable. Herein lies the problem of implementation in large-scale integration. In a system in which the building block is the single circuit, designed and tested to meet rigid specifications, the further testing of an array of circuits cannot be characterized as simple, but it does constitute a tractable problem. In the case of large-scale integration, the individual circuits can no longer be tested. A set of tests must be devised to test whether the chip performs the function for which it is designed, and devising these tests is complicated by the inaccessibility of the terminals of the circuits from which the chip is constructed. If such an array of logic circuitry contains 40 input lines, then the number of possible patterns which can be applied for testing such a function is 2^{40}. Obviously, we cannot blindly present all possible input patterns and test for output patterns, because testing time would be more expensive than all other components of the production costs.

With the foregoing discussion as background, let us now consider a small central processing unit (CPU) of a computer with 10,000 logic circuits; the unit is to be constructed with 100 chips of 100 circuits each. Consider the two extremes: 100 identical chips, or 100 chips that are all different. In the case of the identical chips, one design of layout and test analysis would suffice for the entire unit, whereas in the case of the 100 different chips, 100 designs would be needed. An error on one chip or failure to achieve the design adequately would normally propagate to several chips, so the actual design and verification stage of this central processing unit could be disastrously costly.

The problem of dividing these 10,000 circuits into 100 pieces of 100 circuits

each is called the partitioning problem. The number of different possible partitions of these 10,000 circuits is astronomically large ($\sim 10^{20,000}$). Of this number, one particular partition must be chosen for the actual hardware implementation. The considerations on which these choices are based, such as testability, signal delay paths, possible redundancy of circuits, and design rules which must be included, constitute an area in which a great deal of research is going on in industrial laboratories all over the world. Undoubtedly some of these considerations will force the systems designer to introduce more regularization into the logical structures (*9*). Our investigations have shown that systems as now designed do not lend themselves very readily to the incorporation of repetitive structures, so the 10,000-circuit central processing unit discussed above probably would require at least 60 or more different parts no matter how ingenious the partitioning. Figure 7 shows the results of analyses of the numbers of unique parts required for achieving different levels of integration. We see from these curves that systems as now designed place the designer in a very disadvantageous position for achieving maximum large-scale integration. It is anticipated that, in the future, consideration will be given to the advantages of using repetitive structures so that more favorable curves will be attainable 10 years hence.

Of course, a great deal of work is being directed toward reducing the costs of designing a part. Figure 8 shows a system, now in use, for automatically drawing complex masks, such as those shown in Fig. 6, under computer control. Mask-designing programs exist that require only the feeding of relatively little information to the computer; all the succeeding effort is accomplished automatically. Automatic partitioning, simulation, and test-generation programs are also being developed.

Projections

Where will this revolution in electronics take us? Perhaps the science-fiction writer could deal with this question more comfortably than I can. Certainly, his "wild" projection that computers would design computers and other electronic systems has become a reality, at least in part. We are going to find this application of computers broadening substantially during the next decade. It will become possible to feed rather general requirements for a new electronic system into a computer that will be able to divide the system into manageable parts and then control an automated mask-generating system or direct a photoresist exposure system to implement that system in terms of hardware. Perhaps the instructions will be given to an electron beam apparatus, which can also be used to form the complex patterns required for large-scale integration. We see other techniques being developed, such as beams of dopant atoms with which it may become possible to form the structures needed to perform electronic functions. Research is being directed toward achieving the deposition of high-quality single-crystal silicon on insulating substrates as a means of simplifying the packaging problem. These are really only details. The methods for accomplishing large-scale integration have been developed, so what is required is no longer invention but, rather, the perfecting of existing techniques to the point where high yield of integrated systems can be achieved. How soon these techniques are introduced will be determined by a combination of economic factors which depend on how difficult it will be to implement the techniques and how rapidly their application to electronic systems can be developed.

We will find that the productivity of the individual worker in an electronics factory will climb dramatically—a prediction which implies that electronics, as we know it today, will become very cheap. The net effect will be that electronic techniques will be applied in many areas which as yet have not felt the impact of this new force. Most of us live in homes that have television sets and telephones. The cost of very much more complex electronic systems will become considerably lower than the cost of the family television set. Many of us have wondered what possible use we would have for a computer in our homes. Whenever we are about ready to

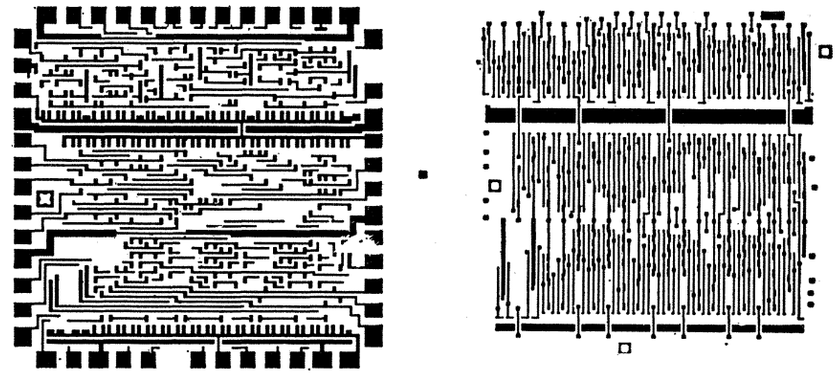
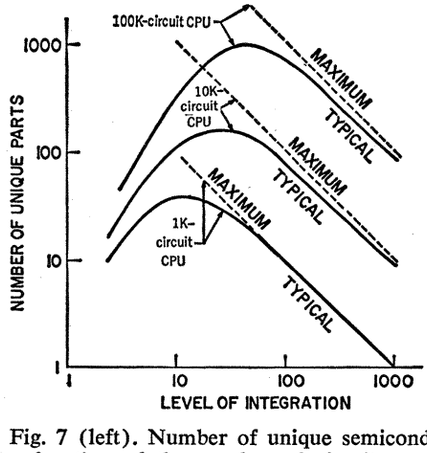


Fig. 6. Diffusion and metalization masks for a 122-circuit logic chip in IGFET technology (see text).



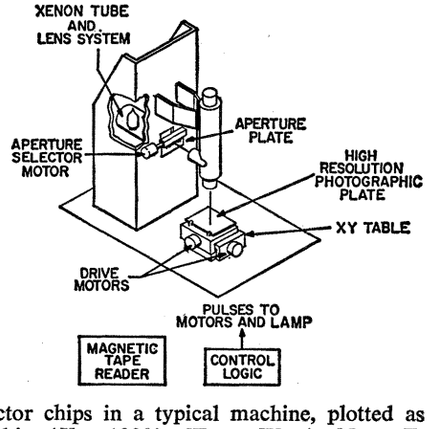


Fig. 7 (left). Number of unique semiconductor chips in a typical machine, plotted as a function of the number of circuits per chip ($K = 1000$). [From W. A. Notz, E. Schischa, J. L. Smith, M. G. Smith, *Electronics* **40**, 130 (1967), with permission] Fig 8. (right). Automatic artwork-generating apparatus for drawing masks such as those shown in Fig. 6.

conclude that we do not need one, we remember judgments made by very competent analysts in the late 1940's and early 1950's to the effect that, if there existed ten computers that could execute instructions at a rate of 1000 per second, then these ten, together, could solve all the problems anyone could possibly imagine. The crux of the matter seems to be that, when an apparently useful device becomes available, the fact of its availability sets a great many more people thinking about the applications than addressed that question during its development. P. E. Haggerty of the Texas Instruments Company remarked, in an address several years ago (*10*), that electronics will become "pervasive." Automated production of large-scale integrated electronics systems is a force making this pervasiveness possible.

Let me list some of the areas in which this "pervasiveness" is now either a fact or an expectation. There is the ubiquitous computer terminal at airline reservation desks. Undergoing rapid development are the terminal in contact with a high-powered computer system for the scientific user; the medical monitoring terminal coupled to the hospital's central computer; management information systems that enable executives to call for digested business data on the basis of which they can make rapid decisions; and, of course, various computer-assisted education programs. We can look forward to the development of computer terminals that will assist doctors in making diagnoses and recording data; to the development of on-line recorders of business transactions, for more effective retail operations; and, eventually, to virtual elimination of the need for money in the form of cash. The reader can extend this list almost indefinitely; the question I contemplate with interest is this: What impact will really inexpensive electronics have on the ingenious toy designers and the generation of young programmers that will follow?

References and Notes

1. N. Nisenoff, *Proc. I.E.E.E. (Inst. Elec. Electron. Engrs.)* **54**, 1820 (1966).
2. E. G. Fubini and M. G. Smith, *Spectrum* **4**, 55 (1967).
3. E. M. Davis, W. E. Harding, R. S. Schwartz, J. J. Corning, *IBM J. Res. Develop.* **8**, 102 (1964).
4. Alundum is a corundum-like material made by fusing alumina in an electric furnace.
5. For a more detailed review of this important technology, see J. W. Lathrop, *Proc. I.E.E.E. (Inst. Elec. Electron. Engrs.)* **52**, 1430 (1964).
6. For a more elaborate discussion of cost model, see B. T. Murphy, *ibid.*, p. 1537.
7. S. Triebwasser, in *Digest of International Solid State Circuits Conference, 1966* (Lewis Winner, New York, 1966), p. 124; M. Canning, R. S. Dunn, G. Jeansonne, *Electronics* **143** (Feb. 1967).
8. R. H. Dennard, private communication.
9. L. C. Hobbs, in *Proceedings of the AFIPS [American Federation of Information Processing Societies] 1966 Fall Joint Computer Conference* (Spartan, Washington, D.C., 1966), pp. 89–96; M. G. Smith and W. A. Notz, *Proceedings of the AFIPS 1967 Fall Joint Computer Conference* (Thompson, Washington, D.C., 1967), pp. 87–94.
10. P. E. Haggerty, *Proc. I.E.E.E. (Inst. Elec. Electron. Engrs.)* **52**, 1400 (1964).
11. I am indebted to many people for very stimulating discussions on the subject of large-scale integration and the revolution in electronics; the individuals who influenced my thinking the most were H. Freitag, D. E. Rosenheim, and M. G. Smith of our laboratory at the IBM Thomas J. Watson Research Center.

Learning of Visceral and Glandular Responses

Recent experiments on animals show the fallacy of an ancient view of the autonomic nervous system.

Neal E. Miller

There is a strong traditional belief in the inferiority of the autonomic nervous system and the visceral responses that it controls. The recent experiments disproving this belief have deep implications for theories of learning, for individual differences in autonomic responses, for the cause and the cure of abnormal psychosomatic symptoms, and possibly also for the understanding of normal homeostasis. Their success encourages investigators to try other unconventional types of training. Before describing these experiments, let me briefly sketch some elements in the history of the deeply entrenched, false belief in the gross inferiority of one major part of the nervous system.

Historical Roots and Modern Ramifications

Since ancient times, reason and the voluntary responses of the skeletal muscles have been considered to be superior, while emotions and the presumably involuntary glandular and visceral responses have been considered to be inferior. This invidious dichotomy appears in the philosophy of Plato (*1*), with his superior rational soul in the head above and inferior souls in the body below. Much later, the great French neuroanatomist Bichat (*2*) distinguished between the cerebrospinal nervous system of the great brain and spinal cord, controlling skeletal responses, and the dual chain of ganglia (which he called "little brains") running down on either side of the spinal cord in the body below and controlling emotional and visceral responses. He indicated his low opinion of the ganglionic system by calling it "vegetative"; he also believed it to be largely independent of the cerebrospinal system, an opinion which is still reflected in our modern name for it, the autonomic nervous system. Considerably later, Cannon (*3*) studied the sympathetic part of the autonomic nervous system and concluded that the different nerves in it all fire simultaneously and are incapable of the finely differentiated individual responses possible for the cerebrospinal system, a conclusion which is enshrined in modern textbooks.

Many, though not all, psychiatrists have made an invidious distinction between the hysterical and other symptoms

The author is professor and head of a Laboratory of Physiological Psychology at the Rockefeller University, New York, New York.