quite unnecessarily to a physical and intellectual ennui.

One way to contemplate the future which does not preempt our choices and which simultaneously puts the computer-man relationship in better perspective is to view the future in terms of four possible scenarios: (i) man without computer, (ii) man with computer, (iii) computers without man, and (iv) computers against man.

Within the setting of these scenarios, what we know about the evolution of computers to date leads me to predict that:

1) Progress, or changes, in the advanced, imaginative uses of computers will be despairingly slow—certainly much slower than in the first 25 years of computer development.

2) The decreasing costs and decreasing size of computers and logical devices will put these scientific artifacts into the hands of large numbers of individuals. We will see spurts of that "basement creativity" for which Americans are so renowned. Computer-related advances will be many, random, and beneficial. However, because of constraints to progress, these advances will be localized without large-scale diffusion.

3) Man coupled with computers will outlast man without computers. This surmise appears reasonable for both individuals and groups in society.

4) Man will continue to increase the number of "intelligent" tasks for computers faster than he does for himself. This paradoxical situation prevails because (i) we like to coax the utmost out of our most fascinating invention; (ii) we are intellectually lazy; and (iii) many people already feel insecure without the computer and are, in short, productively addicted to it.

5) Computers will provide to the individual more control over his personal environment than he has ever before been able to exercise. This capability will result from the miniaturization of computer components along with the decreased cost of computer hardware. Controls now possible range from individual self-paced instruction to the monitoring in real time of vital health signs, to surveillance for public protection of dangerous surroundings such as alleys and hallways. Attainment of these kinds of individual control will be accomplished principally through local ingenuity.

6) Major efforts will be directed toward the use of computers for increasing public accountability. This will take the form of more computers used for more record-keeping tasks.

7) In spite of all man-made constraints, there will be an irreversible but slow trek to realize with computers, forms of intelligent behavior that are essentially limitless, transcending man and computer taken separately.

I feel quite sanguine that people will not, as is so frequently stated, become the victim rather than the master of computers. Further, we are all most fortunate to be part of the process that will be making monumentally important decisions about man's intellectual future.

References and Notes

1. H. D. Huskey and V. R. Huskey, *IEEE Trans. Comput.* **C-25** (2), 1190 (1976).
2. G. K. Kostopoulos, *Digital Methods*, (Wiley Interscience, New York, 1975), p. 1.
3. G. Birkhoff, "Applied mathematics and its future," speech at National Bureau of Standards, 26 March 1976.
4. A. W. Burks, H. H. Goldstine, J. von Neumann, *Preliminary Discussion of the Logical Design on an Electronic Computing Instrument* (Institute for Advanced Study, Princeton, N.J., 1946), vol. 1, part 1.
5. C. Eames and R. Eames, *A Computer Perspective* (Harvard Univ. Press, Cambridge, Mass., 1900), pp. 6 and 128–131.
6. D. A. Bell, *Fundamentals of Electronic Devices* (Reston Publishing, Reston, Va., 1975), pp. 1–7.
7. L. M. Branscomb, *Phys. Today* **29** (1), 55 (1976).
8. *American National Dictionary for Information Processing* (American National Standards Institute, New York, in press).
9. I. W. Cotton, "Executive Guide to Computer Networking," National Bureau of Standards, Washington, D.C., publication in process (available in draft).
10. Information supplied by DATAPRO Research Corporation, Delran, N.J., 1976.
11. D. A. Hodges, *Comput. Des.* **15** (2), 77 (1976).
12. R. M. Davis, *Computer* **6** (4), 14 (1973).
13. M. Greenberger, M. A. Crenson, B. L. Crissey, *Models in the Policy Process* (Russell Sage Foundation New York, 1976), pp. 1–42.
14. Data supplied by the International Data Corporation, Waltham, Mass., 1976.
15. General Services Administration, Inventory of Automatic Data Processing Equipment in the U.S. Government, 30 September 1976.
16. See H. H. Goldstine, *The Computer from Pascal to von Neumann* (Princeton Univ. Press, Princeton, N.J., 1972), pp. 266–270.
17. R. T. Yeh, *Computer* **8** (5), 15 (1975).
18. R. London, *Proc. Assoc. Comput. Mach. Annu. Conf.* **1**, 39 (1972).
19. United States versus IBM, plaintiff's pretrial brief 1 of 17 October 1974, Southern District Court, New York.
20. N. Wiener, *Cybernetics* (MIT Press, Cambridge, Mass., 1948), pp. 13–15.
21. C. R. Spitzer, *Spectrum* **13** (10), 92 (1976).

Large-Scale Integration: What Is Yet to Come?

Robert N. Noyce

A single large-scale integrated circuit in 1977 can contain more active elements than the most complex electronic equipment ever built 25 years ago. Enormous advances have been made in the last two decades, and we may ask whether this technology is reaching a limit or whether it will continue to advance at a breathtaking pace. By their nature, integrated circuits are small and are limited in power dissipation. As a result, their primary use is in information manipulation, not in areas where high power is required, such as in transmission or in servo drivers. However, a scanning of the titles of this issue shows that information-handling or computerlike devices are central to electronics today. Thus, large-scale integration can have a major influence on the direction of electronics in the future.

The most obvious effect of the development of integrated circuits has been to reduce the size of electronic equipment. However, the principal advantages that have accrued lie in the reduction of cost and in the improvements in performance, reproducibility, maintainability, and reliability.

Interconnections Are the Problem

Interconnections have been, and are, the major problem in computer hardware. Because of the physical size of these interconnections, delays are incurred in transmitting signals from origin to destination with the computer. Because of the capacitance of the interconnections, a power loss is incurred every time a signal is imposed on the

The author is chairman of the Intel Corporation, Santa Clara, California 95051.

This article was adapted from a forthcoming book on the future impact of computers and information processing, M. L. Dertouzos and J. Moses, Eds., MIT Press.

line, necessitating high-power drivers in the system which are not needed to drive the actual logic of the computer. And because the interconnections themselves act as antennas, they transmit to and pick up from other interconnections unwanted signals, or noise. As a result, there is an exploding design of increasing power, increasing noise, and increasing size, which requires higher power, longer delays, and higher cost. With current technology, the essential logic and memory of a large modern computer could be contained within 10 to 100 cubic centimeters of silicon chips. Interconnections add not only to the size but to the apparent complexity of the large computer, resulting in maintenance, documentation, and reliability problems that increase in severity with size. The integrated circuit is the component industry's solution to the interconnection problems.

If we consider only the cost of the active elements in the integrated circuit, we quickly conclude that the single transistor is the most cost-effective.

A simple model would say that if a particular slice of silicon yielded chips of which only 10 percent were good because of the inclusion of random defects, then if it were used to make chips of twice the area it would yield only 1 percent good chips. (Actually the situation is not this bad, since defects are not randomly distributed.) The cost for twice the function then would be 20 times as great (twice the processing cost since twice the area of silicon is used, and ten times the cost because of loss of yield). Clearly, if the cost of the active silicon were dominant, such a doubling of complexity would not be cost-effective; carried to an extreme, the single transistor would be the most cost-effective.

The other major cost borne by the component manufacturer is that of assembling and testing the devices. Assembly is the process of putting the tiny silicon active element in a housing, which includes a mechanical transition from the microscopic interconnections of the integrated circuit to the sizes normally encountered in electronic equipment—that is, from a lead separation of 10 micrometers to one of 2500 μm or 2.5 millimeters. As a first approximation, assembly costs are independent of the function included on the integrated circuit chip, although they will increase somewhat as the number of electrical connections to large chips increases. Similarly, test costs increase much more slowly than the complexity of the chip being tested, although very sophisticated test equipment is required to achieve this result. Thus, the total cost per function will be made up of two elements: one increasing with complexity, of the general form ae^{bN}, representing the cost of the silicon chip; and the other, of the form c/N, representing the cost of assembly and test, where N is the number of functions included. This cost will have a minimum, as indicated in Fig. 1. As processes for manufacturing integrated circuits have been perfected and yields of good circuits have been improved, this minimum cost point has moved to circuits of higher complexity. It has been the strategy of the semiconductor device manufacturers to supply circuits that are near this minimum at any given time.

An examination of the integrated circuits offered by the industry as a function of time provides an estimate of how the complexity of products at the minimum cost point has increased with time, even though such products are introduced before they represent the lowest cost per function point. As indicated in Fig. 2, the complexity of products at the minimum cost point has doubled every year since the introduction of the integrated circuit. If the present rate of increase of complexity were to continue, integrated circuits with 10^9 elements would be available in 20 years.

Higher levels of integration have been achieved in three ways: by increasing practical chip size through the reduction of random defects, by introducing circuit innovations that have allowed higher functional density, and by making the individual circuit elements smaller. The last has been the most important factor. We can expect the achievement of smaller dimensions to be the principal contributor to increasing levels of integration in the future. As device dimensions are made smaller, all of the device parameters move in a favorable direction, as can be seen by scaling the device dimensions and noting the changes in these parameters.

The electric fields cannot be increased as the dimensions are changed, since devices are designed with limiting electric fields now, and higher fields will cause avalanche breakdown or quantum mechanical tunneling. Thus, if x is the characteristic dimension of the device, the operating voltages of the device will vary as

$$V \propto x$$

Since the fields terminate in sheets of charge with finite thickness, the charge densities must increase to scale this

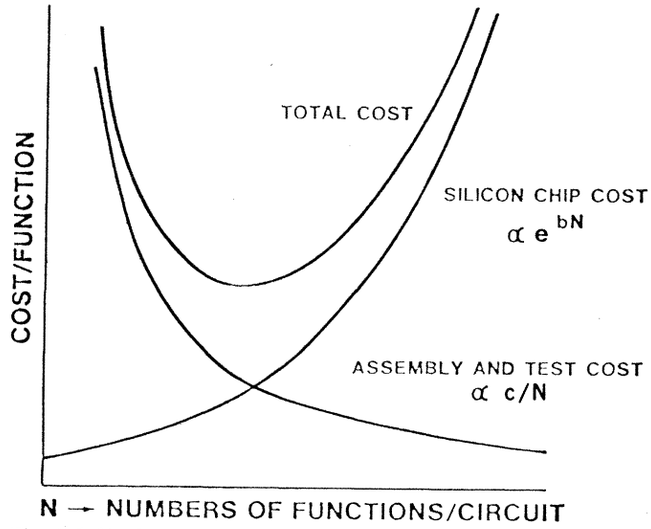


Fig. 1 (left). Cost per function versus circuit complexity.

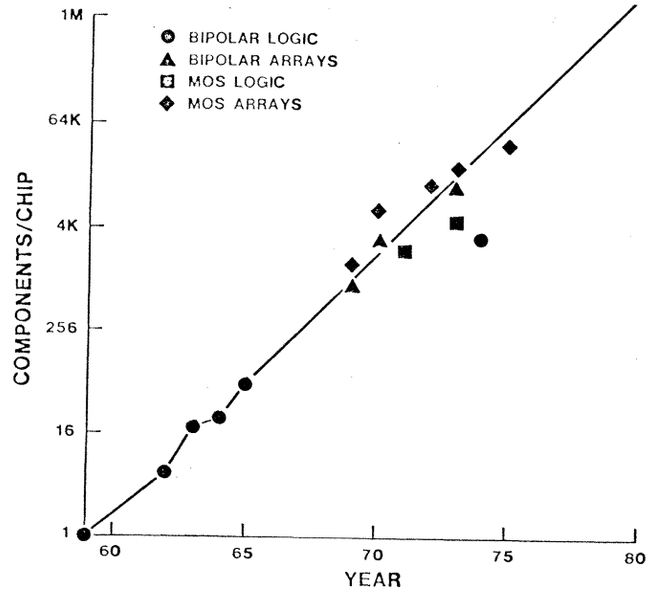


Fig. 2 (right). Circuit complexity versus time of introduction.

thickness. Thus the impurity densities in the device vary inversely with x. Similarly, the mobile charge n must increase as $1/x$. The area A of the conducting path varies with x^2, so the current varies as

$$I \propto nA \propto x$$

since the field is constant. The characteristic impedance of the circuit is constant

$$Z \propto V/I = \text{constant}$$

We may look at circuit delay times from two points of view, in terms of carrier transit time, τ_t, and capacitance charging time, τ_c. The transit time τ_t will be proportional to the distance traveled

$$\tau_t \propto x$$

since the field is constant. Charging the collector or drain capacitance depends on the voltage, current, and capacitance. The capacitance, C, varies as the dimensions, so

$$\tau_c = \frac{CV}{I} \propto x$$

which is similar to the variation in transit time.

The power per circuit will be proportional to the product of current and voltages

$$P \propto VI \propto x^2$$

The power density, P/x^2, is then constant, since the area of the circuit varies as x^2.

Thus, as dimensions are made smaller, all of these variables move in a favorable, or at least not unfavorable, direction. Speed increases, density increases, and power density does not increase.

Since the operating voltages decrease as the dimensions decrease, the allowable voltage drop along interconnections in the integrated circuit will decrease in proportion. A scaling argument, however, shows that these voltage drops will remain constant. For the normal metal interconnections, today's designs are far from the limits at which problems would arise. Some interconnection schemes, such as the use of polycrystalline silicon, are near the limit now. However, the use of other techniques does not appear to pose a great problem.

Figure 3 shows the historical trend in minimum average dimensions in advanced integrated circuits. Reduction by a factor of 2 is being achieved every 5 years. Were this to continue, circuit densities would increase by a factor of 64 and speeds would increase by a factor of 8 in the next 15 years.

Defect Density

The production of integrated circuits has always been a "yield" problem. Many circuits are made, the defective ones are thrown away, and the good ones are sold. Defects can arise from many sources. The photomasks used may have pinholes in dark areas, or opaque specks in areas that should be clear. Severe defects in the basic silicon crystal can make the circuit inoperative. Dust in the photoprinting operation or other processing steps that affects a critical spot in the circuit will cause failures. Misalignment in successive photoengraving steps or lack of control of critical dimensions and impurity concentrations will make the circuit inoperative.

The progress in reducing defect densities can be seen from the history of silicon chip sizes of new products, as shown in Fig. 4. The absolute defect density is approximately the inverse of the chip size, since at higher defect densities the chip cost would be prohibitively high, and at lower defect densities higher levels of integration give a cost advantage.

In 15 years the extrapolated reduction in dimensions will reduce internal gate delays to the order of 1.0 nanosecond. Propagation delays will then begin to become significant if system dimensions are much more than 10 cm. During the same period, the silicon chip dimensions can be expected to increase to about 2.5 cm on a side. Thus, delays due to on-chip propagation should not be a problem, but delays in signal propagation between chips will be a significant factor.

The combined density and chip size extrapolations would result in an increase of functional complexity by a factor of over 2000 during the next 15 years. Costs would increase only slowly from those of today's complex integrated circuits, resulting in a cost reduction of 100 to 1 or even 1000 to 1. Clearly there is a motivation to try to achieve this result.

Practical Limits

Since most production technologies use photolithography to define the circuits, the wavelength of light looms large as a limit to the dimensions that can be achieved. The resolution of the printed pattern is involved, but a more difficult problem is the reindexing of successive photolithographic steps necessary to build up the integrated circuit structure. Today's techniques require that the previously printed pattern be examined (either by human eye or by automatic equipment) and aligned with the new pattern to be printed. This examination is now made with visible light, with the resolution which that implies. Realignment more accurate than about three wavelengths of light is not yet practical. The

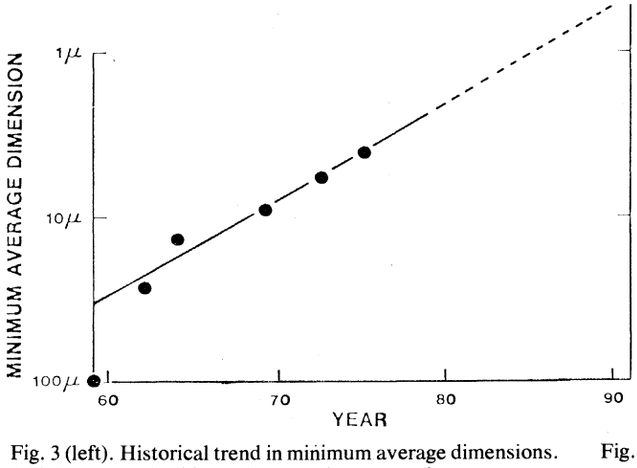


Fig. 3 (left). Historical trend in minimum average dimensions.

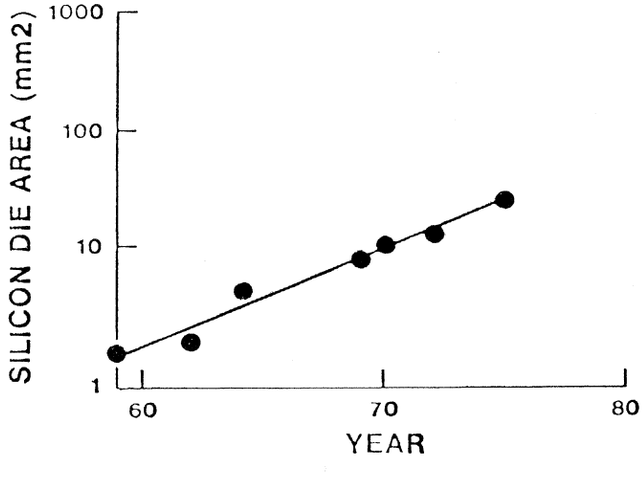


Fig. 4 (right). Trend in silicon chip sizes in new products.

allowance for this amount of misalignment in the design of the circuit means that much smaller circuits cannot be produced by this technique.

One approach to avoiding this limitation is to remove the need for accurate alignment by using "self-aligned" techniques. For example, in the silicon gate technique used in metal oxide semiconductor (MOS) circuits the gate itself and the source-drain spacing are defined in the same photolithographic step. The use of self-aligned techniques eliminates the allowance for misalignment of some masking steps from the design of the circuits and allows tighter packing densities.

Another approach is to develop new device forms in which the critical dimensions are defined by techniques other than photoengraving. An example is the diffused MOS transistor, in which the source-drain spacing is defined by solid-state diffusion rather than by photolithography.

A third approach to overcoming resolution limitations is to use electron beams rather than optical techniques. Indeed, the use of electron beams for the preparation of high-resolution photomasks is now being introduced (*1*). However, production use of electron beam lithography in printing on the silicon wafers appears to be 5 to 10 years away. The many problems of inspection and realignment techniques must be solved. In addition, production-worthy equipment must be developed. The economic return on the use of the electron beam is not clear, particularly for the pioneer who assumes most of the cost of developing the technology. Therefore, the most likely scenario is that the electron beam will be used to prepare photomasks, but optical printing techniques will be used on the silicon.

There appears to be no fundamental limitation to the reduction of defect densities. The integrated circuit industry has proved very resourceful in solving these problems. Progress has been steady and seems to be determined by the amount of effort that is expended to perfect the production processes. Since the economic return is very high, we can expect the effort to continue, with results similar to those experienced in the past.

Innovation in circuit or device forms to improve densities seems to be reaching a limit, perhaps as a result of our inability to anticipate future innovations in new circuit forms. Possibilities for improvement may come from using three-dimensional forms rather than the present two-dimensional technology.

Redundancy in some form would allow much higher levels of integration than would be possible otherwise. It appears that as circuits become more complex, the overhead expense for achieving fault-tolerant circuits is a smaller proportion of the total. For example, error correction for an 8-bit word requires an additional 4 bits, whereas correction for a 64-bit word requires only 7 additional bits. A similar case can probably be made for logic. The relative cost of including redundancy is less in more complex systems. Today error correction is used in large memory systems to reduce failure rates. Partially good memory components are used in some systems by disabling faulty addresses. As new schemes are developed for the use of faulty circuits, the permissible chip sizes could expand greatly, allowing significant cost reduction with higher levels of integration.

What Is Still to Come?

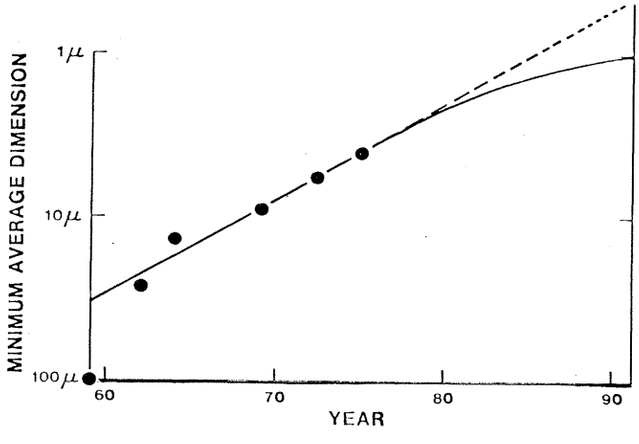
The drive to higher levels of integration has great momentum and will continue. Minimum dimensions will continue to decrease, but at a decreasing rate as optical limits are approached, reaching 2 μm in 5 years and 1 μm in 15 years (Fig. 5). Scaling arguments show that speed should then increase by a factor of 4 by 1991. If defect densities are reduced as they have been in the past, chip sizes will increase by a factor of 3 in 5 years and by a factor of 25 in 15 years (Fig. 6).

As a result of these factors, components providing 65 to 131 kilobits of memory with an access time of 100 nsec should appear in 1981 and the megabit memory chip (2^{20} bits) should appear 10 years later. The use of redundancy could accelerate these times. Component costs should be comparable to those of today's memory components, or 10^{-3} cent per bit.

For noniterated circuits such as control logic circuits, the level of integration will be lower because of the inefficiencies inherent in packing random logic. However, the level that could be achieved in 5 years would be approximately 25,000 gates, and in 10 years about 250,000 gates. These numbers exceed the gate counts of today's medium and large processors, respectively. Internal gate delays of these systems would be comparable to those of today's high-speed computers. If the amortization of design costs could be neglected, such logic arrays could be produced for less than $100 or a cost of less than 0.4 cent per gate in 5 years and less than 0.04 cent per gate in 15 years (Fig. 7).

It is still questionable whether such high levels of integration of logic circuits will be attempted. The industry has already achieved the complexity needed for today's mass markets—for example, in the digital watch or the calculator—and the profit potential that motivates these developments can only be achieved by production in high volume.

The basic technology will be developed, however, to supply the demand for



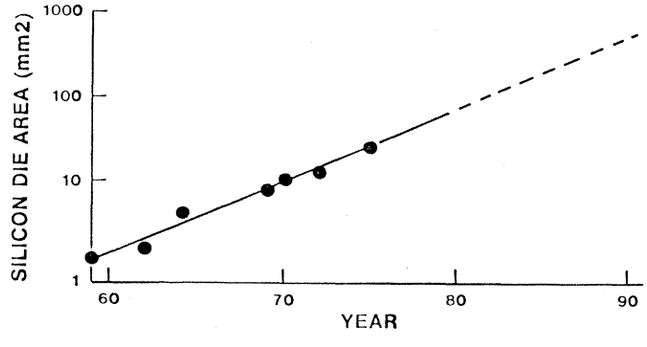


Fig. 5 (left). Projected minimum average dimension in advanced integrated circuits. Fig. 6 (right). Projected trend in silicon chip sizes.

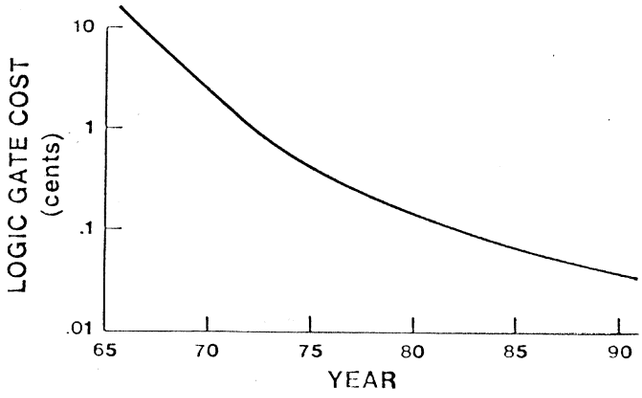


Fig. 7. Cost of logic arrays versus time.

memory. Fortunately perhaps, there is an insatiable demand for memory, and the market is completely elastic at today's usage levels. The question, then, becomes one of the cost of design of high-complexity logic for a particular application. At high levels of integration, many more unique circuit designs are necessary than for Boolean or simple logic functions. The market for each unique circuit becomes smaller, while the design costs become much higher.

The integrated circuit industry faced this problem with the introduction of large-scale integration and found a solution in the concept of the microprocessor. The microprocessor was conceived as a method of supplying many diverse requirements of random logic, thus allowing one or a small number of circuit designs to achieve a high enough market to justify their development. The true impact of the microprocessor is just being seen today, but even now it is having a radical influence on our thinking about future computing systems. We may expect the capability of the microprocessor to increase dramatically in the future, if for no other reason than that it can be done. The impact of the microprocessor is first being felt in smaller computing systems, such as terminals, controllers, and video games. As its capabilities increase, the equivalents of today's processors will be built using this concept.

Several changes will probably occur in the computer industry as a result of the microprocessor. Standardization about a small number of microprocessors will be forced by economics, since manufacturing them in volume will reduce costs substantially. It now appears that identical microprocessors will generally be available from many sources. This will not only assure competitive prices but also make possible the development of software sources independent of the hardware manufacturers. The microprocessor will become another building block, like a flip-flop or an adder, for the construction of computing systems. This suggests that architecture will change to take advantage of these very cost-effective building blocks. Increasing throughput by using identical blocks in parallel should be favored, and networks of semi-independent processors should be more cost-effective than large time-shared machines.

Structural changes will occur as well, since the microcomputer designer must consider the economics of component production and the economics of system use simultaneously. This suggests that there may be some amalgamation of the components and system producers, or at least some mutual penetration of these industries.

The potential for developing inexpensive processing power is truly awesome. With low-cost processing available, many new tasks may be undertaken which are totally uneconomical today. In addition, with more powerful processing the requirements for input-output devices may be reduced to simpler forms of information transfer such as audio input and response. Our major challenge is likely to be defining the tasks to be accomplished with this technology, and developing the software with which to accomplish those tasks.

Summary

The rapid development of large-scale integrated circuits in the last two decades has revolutionized information handling. Higher levels of integration have been achieved principally by making the individual circuit elements smaller, but reduction of random defects and innovations in circuit design have also been important factors. Practical limits to the size of integrated circuits, such as those imposed by the use of photolithography to define the circuits, can be avoided by using other methods, such as solid-state diffusion or electron beams.

The prospects are that dimensions and defect densities will continue to be reduced while speed increases. Redundancy in circuits will lead to lower costs and higher levels of integration. As logic becomes more complex and applications narrower, the increased number of unique circuit designs needed tends to increase costs. The microprocessor was developed to answer this problem and will have a radical influence on future computing systems.

References

1. R. W. Keyes [*Science* **195**, 1230 (1977)] discusses the practical limits to resolution with electron beams.