1 The New Technology

Here Comes the Second Computer Revolution

Gene Bylinsky

The first full-length feature on the microelectronics revolution to appear in a nonspecialist publication, this article was published in the US *magazine* Fortune *in November 1975. Others picked up the story (see the guide to further reading), but it was not until 1978 that the silicon chip became big news.*

Less than thirty years ago, electrical engineer J. Presper Eckert Jr. and physicist John W. Mauchly, at times assisted by as many as fifty helpers, laboriously built the world's first electronic digital computer. Their ENIAC (Electronic Numerical Integrator and Computer) was a fickle monster that weighed thirty tons and ran on 18,000 vacuum tubes – when it ran. But it started the computer revolution.

Now under way is a new expansion of electronics into our lives, a second computer revolution that will transform ordinary products and create many new ones. The instrument of change is an electronic data-processing machine so tiny that it could easily have been lost in the pocket of one of those ENIAC tubes. This remarkable device is the microcomputer, also known as the computer-on-a-chip. In its basic configuration, it consists of just that – a complex of circuits on a chip of silicon about the size of the first three letters in the word ENIAC as printed here. Yet even a medium-strength microcomputer can perform 100,000 calculations a second, twenty times as many as ENIAC could.

This smallest of all data-processing machines was invented six years ago, but its mass applications are just beginning to explode, setting off reverberations that will affect work and play, the profitability and productivity of corporations, and the nature of the computer industry itself. For the microcomputer provides an awesome amount of computer power in a package that in its simplest form costs less than $10 bought in quantity and easily fits inside a matchbox. Accessory devices bring microcomputer prices to between $50 and $250 apiece, to be sure, but that's still a lot less than the thousands of dollars in minicomputer costs.

And unlike the familiar older computers that come in their own boxes, the microcomputer is mounted on a small board that can be made to fit easily and unobtrusively into a corner of an electric typewriter, a butcher's scale, a cash register, a microwave oven, a gas pump, a traffic light, a complex scientific instrument such as a gas chromatograph, and any of a myriad other devices whose capabilities already are being enhanced by these slices of electronic brainpower. Soon microcomputers will start replacing wheels, gears, and mechanical relays in a wide variety of control applications, because it's much more efficient to move electrons around than mechanical parts.

To cite these applications and capabilities, as well as many other uses to come in the home, the factory, and the automobile, is to do only pale justice to this marvellous invention. What sets any computer apart from every other kind of machine is its stored and alterable program, which allows one computer to perform many different tasks in response to simple program changes. Now the microcomputer can impart this power, in a compact form and at a low price, to many other machines and devices.

In the most common form of microcomputer, furthermore, a user can change the program simply by unplugging a tiny memory chip and putting a new one in its place. To show off this versatility, Pro-Log Corp. of Monterey, California, built a demonstration apparatus that in its original version is a digital clock; when a program chip that runs the clock is removed and another is put in its place, the thing suddenly starts belting out a tinny version of the theme from *The Sting*. With still another memory chip, it becomes a rudimentary piano.

Besides providing versatility for users, the microcomputer makes possible large economies in manufacturing. Now a manufacturer can buy a standard microcomputer system for many different products and use a different program chip with each. By doing so, the manufacturer can save substantial amounts of money since a single microcomputer can replace as many as 200 individual logic chips, which cost about $3 each.

The use of microcomputers, moreover, can substantially reduce service and warranty costs because the reliability of the electronic portion of a device is increased up to tenfold. A microcomputer that replaces, say, fifty integrated circuits does away with about 1800 interconnections – where most failures occur in electronics. The microcomputer, in other words, is one of those rare innovations that at the same time reduce the cost of manufacturing *and* enhance the capabilities and value of the product. Thus the microcomputer may be the best technological antidote for inflation in quite a while.

Even the men who make and use microcomputers say that they haven't yet grasped the device's full implications, but they know the implications are large and far-reaching. Fairly typical is the comment of Edward L. Gelbach, senior vice president at Intel Corp., the Santa Clara, California, semiconductor company where the tiny computer was invented. "The microcomputer," he says, "is almost too good to be true."

The microcomputer is the logical end result of the electronics industry's headlong drive to miniaturize. The industry has galloped through three generations of components in as many decades. In the late 1950s, the transistor replaced the vacuum tube. Within a few years the transistor itself gave way to "large-scale integration", or LSI, the technique that now places thousands of micro-miniaturized transistors – an integrated circuit – on a sliver of silicon only a fraction of an inch thick. LSI made possible the suitcase-sized minicomputer.

The semiconductor logic circuit, of course, contained the seed of the microcomputer, since the chip had logic elements on it – the transistors. But the individual chips were designed to perform limited tasks. Accordingly, the central processing units of large computers were made up of hundreds, or thousands, of integrated circuits.

Logic chips were also employed for control or arithmetic functions in specialized applications. In what became known as "hardwired logic" systems, chips and other individual components were soldered into a rigid pattern on a so-called printed-circuit board. The fixed interconnections served as the program. Curiously, it was even less flexible than ENIAC's primitive array of plug-in wires that could be moved around to change the program.

The electronic calculator, in all but the latest versions, uses hardwired logic. The arithmetic functions, or the operating program instructions, are embedded in the chips, while the application program is in the user's head – his instructions yield the desired calculations.

A young Intel engineer, M. E. Hoff Jr, envisaged a different way of employing the new electronic capabilities. He had received a Ph.D. in electronics from Stanford University, where he had become accustomed to solving problems with general purpose data-processing machines. In 1969 he found himself in charge of a project that Intel took on for Busicom, a Japanese calculator company. Busicom wanted Intel to produce calculator chips of Japanese design. The logic circuits were spread around eleven chips and the complexity of the design would have taxed Intel's capabilities – it was then a small company.

Hoff saw a way to improve on the Japanese design by making a bold technological leap. Intel had pioneered in the development of semiconductor memory chips to be used in large computers. (See "How Intel Won Its Bet on Memory Chips", *Fortune*, November 1973.) In the intricate innards of a memory chip, Hoff knew, it was possible to store a program to run a minuscule computing circuit.

In his preliminary design, Hoff condensed the layout onto three chips. He put the computer's "brain", its central processing unit, on a single chip of silicon. That was possible because the semiconductor industry had developed a means of inscribing very complex circuits on tiny surfaces. A master drawing, usually 500 times as large as the actual chip, is reduced photographically to microminiature size. The photo images are then transferred to the chip by a technique similar to photoengraving.

Hoff's CPU on a chip became known as the microprocessor. To the microprocessor, he attached two memory chips, one to move data in and out of the CPU and one to provide the program to drive the CPU. Hoff now had in hand a rudimentary general-purpose computer that not only could run a complex calculator but also could control an elevator or a set of traffic lights, and perform many other tasks, depending on its program. The microcomputer was slower than minicomputers, but it could be mass-produced as a component, on the same high-volume lines where Intel made memory chips – a surprising development that would suddenly put the semiconductor company into the computer business.

Hoff had strong backers in Intel's top executives: President Gordon E. Moore and Chairman Robert N. Noyce, the co-inventor of the integrated circuit. Unlike many other specialists, Noyce and Moore had sensed the potential of the microcomputer early on, and they lent enthusiastic support to Hoff's project. Most others had visualized a computer-on-a-chip as being something extremely expensive and far in the future. When in the late 1960s Noyce suggested at a conference that the next decade would see the development of a computer-on-a-chip, one of his fellow panellists typically remarked in all seriousness: "Gee, I certainly wouldn't want to lose my whole computer through a crack in the floor." Noyce told the man: "You have it all wrong, because you'll have 100 more sitting on your desk, so it won't matter if you lose one."

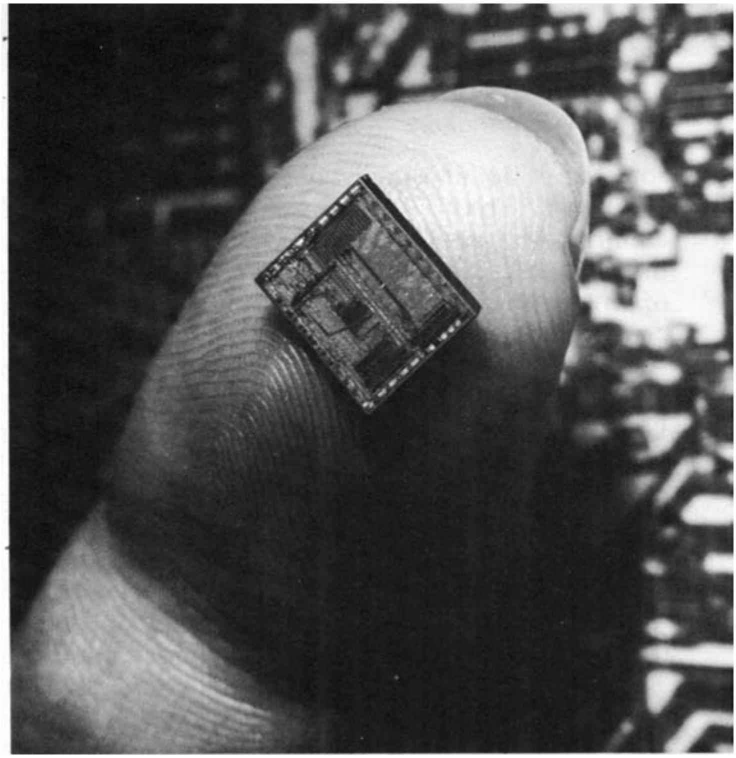
After other Intel engineers who took over the detailed design work got through with it, Hoff's invention contained 2250 microminiaturized transistors on a chip slightly less than one sixth of an inch long and one eighth of an inch wide, and each of those microscopic transistors was roughly equal to an ENIAC vacuum tube. Intel labelled the microprocessor chip 4004, and the whole microcomputer MCS-4 (microcomputer system 4). "The 4004 will probably be as famous as the ENIAC," says an admiring Motorola executive. Despite its small size, the 4004 just about matched ENIAC's computational power. It also matched the capability of an IBM machine of the early 1960s that sold for $30,000 and whose central processing unit took up the space of an office desk. If any had suggested in the days of ENIAC that this kind of advance would take place so soon, says Presper Eckert, now a vice-president at Sperry Univac, the idea would have struck him as "outlandish".

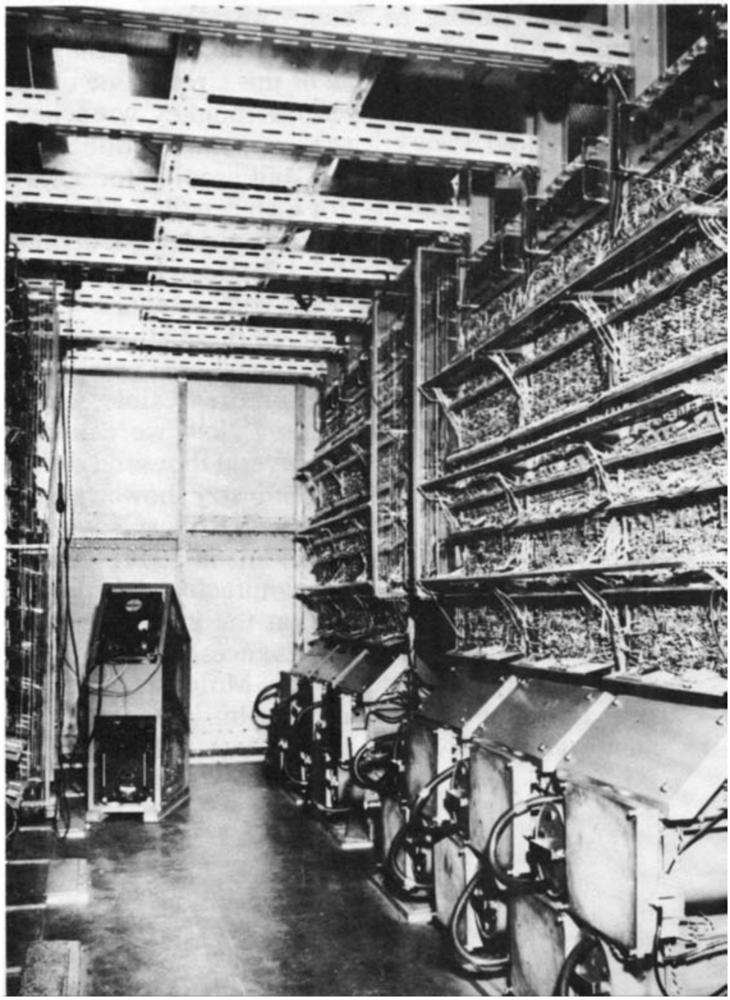
For logic and systems designers the appearance of the microcomputer brought with it a dramatic change in the way they employed electronics. They could now replace all those rigid hardwired logic systems with microcomputers, because they could store program sequences in the labyrinthine circuits of the memory chip instead of using individual logic chips and discrete components to implement the program. Engineers thus could substitute program code words for hardware parts.

For the semiconductor industry the arrival of the microprocessor on a chip signalled the end of a costly search for ways to reduce the complicated technology to more generalized applications. "The problem," says Moore of Intel, "was that as the technology got more complex you couldn't find any generality to the circuit functions. What customers wanted was one of this circuit, one of that circuit, to build a system." Such demands threw monkey wrenches into the industry's efforts to hold down costs through mass production.

The industry kept flailing and groping for ways to master the problem. Texas Instruments, for instance, had a big project aimed at using computer-guided design to make production of integrated logic components more flexible. Fairchild Semiconductor talked about turning out as many as 500 different logic components a week to suit the requirements of different customers. In these attempts, engineers were trying to force the technology to become more flexible. Ted Hoff's solution, to make the internal design itself more flexible, was far more elegant and more powerful. Says Moore: "Now we can make a single microprocessor chip and sell it for several thousand different applications."

At first the semiconductor industry showed surprisingly little interest in this great leap in its technology. Robert Noyce recalls that when Intel introduced the microcomputer late in 1971, the industry's reaction was "ho hum". Semiconductor manufacturers had made so many extravagant promises in the past that the industry seemed to have become immune to claims of real advances. Besides, the big semiconductor companies – Texas Instruments, Motorola, and Fairchild – were preoccupied with their large current business, integrated circuits and calculator chips. "Looking back," says J. Fred Bucy, TI's executive vice-





Ted Hoff's miracle chip shown here is a little more convenient in shape and size than TREAC, an equivalent 1953 computer developed by Britain's Telecommunications Research Establishment.
Above courtesy of Intel Corporation; below courtesy of Professor A. M. Uttley

president and chief operating officer, "we probably should have started on microcomputers earlier."

Only Rockwell International and National Semiconductor got into the field early on, about a year after Intel. Fairchild came out with a microprocessor chip that it sold primarily to calculator manufacturers. It took another six months or so before the new economics of the microcomputer stung the other giants into action. By that time, hardly anyone could have missed the message: a microprocessor and its memory could replace a lot of individual logic chips – anywhere from ten to 200. To speed the adoption of microcomputers, Intel undertook to recast the thinking of industrial-design engineers – the company taught 5000 engineers the use of the microcomputer in the early 1970s and another 5000 or so later on. Once these engineers started ordering the tiny computers in some quantity, the big companies, as Noyce puts it, said: "We've got to get on board here."

They rushed to get on board by "second-sourcing" – that is copying – Intel's microcomputers. Second-sourcing is a common practice in the semiconductor industry. More often than not, it is done without the original manufacturer's permission or cooperation, but the practice is nonetheless widely accepted by the companies involved. It works to the benefit of the user in establishing a competitive source for the component as well as a backup for the original manufacturer. In fact, users normally demand second-sourcing.

Second-sourcing microcomputers proved to be a complex task, however. What's more, Intel kept moving. It followed up the 4004 with a more capacious 8008 model in 1972, and towards the end of 1973 brought out its second-generation microcomputer, the 8080. This was twenty times faster than the 4004. Even then most competitors had no microcomputers of their own to offer. The first real competition to the 8080 was Motorola's 6800, which came a year afterward. The late starters began to catch up this year (1975) when Texas Instruments, General Instrument, and others announced microcomputer models of their own. TI also introduced its copy of the 8080.

To paper over the gap, some nimble competitors upgraded calculator chips and started calling them computers-on-a-chip. With memory on the same square of silicon, these basic units can perform simple and even medium-complexity control functions – running washing machines or microwave ovens, for instance. TI, Rockwell, and others now offer such chips. The TI product, TMS 1000, sells for as little as $4 in large quantities.

All these companies, and many others, are battling for a market that so far is fairly small – this year it will amount to only about $50m. But it is expected to expand to $150m next year, and to reach $450m by 1980.

In these estimates, the microprocessor chips account for only fifteen to twenty per cent of the dollar total, with memories and other components making up the bulk of the new business.

Applications of microcomputers today are tilted heavily towards data-processing equipment of various kinds, including computer terminals and other accessories. The other major market is retailing equipment – electronic cash registers and point-of-sale terminals. But the picture is expected to change drastically in a few years as microcomputers invade consumer products in force. TI estimates that consumer product uses will account for about one third of the predicted $450m-a-year market for microprocessors in 1980.

In their capabilities, microcomputers cover quite a range of applications. A simple microcomputer can act as a miniature controller, replacing an electromechanical relay or hardwired logic systems. A more powerful model, such as the 8080, can control a computer printer, or a whole series of them. Still more powerful models begin to match – and some already exceed – minicomputers in their computational speeds.

The tiny computer is beginning to generate not only new products but new companies as well. Says Gordon Hoffman, an executive at Mostek, a Dallas semiconductor house: "A lot of big companies are going to be improperly prepared to take advantage of the microcomputer. If they don't take advantage of it, they may find themselves out in the cold when a little upstart comes along and says: 'I can do it better with a microcomputer.'"

That kind of competition has already begun, with many fast-moving small companies taking advantage of the microcomputer's mighty power. A few examples:

> Chemetrics Corp. of Burlingame, California, only two years old, has brought out an advanced blood-chemistry analyser.
>
> Electro Units Corp. of San Jose has developed an electronic control system for bars; it doles out precisely measured drinks and serves as an attentive inventory controller too.
>
> Telesensory System Inc. of Palo Alto is introducing this autumn a "talking" calculator for the blind, with a recorded vocabulary of twenty-four words for spoken verification of calculation steps and results.

Large companies, of course, are also using the capabilities of the computer-on-a-chip to turn out new products. Among them:

> General Electric, which is looking into many possible applications, recently introduced a robot industrial tool run by a tiny computer.
>
> AMF, with the aid of Motorola, developed an automatic scorer now being demonstrated in bowling alleys.

> Tappan Co. is designing a microwave oven with "touch-and-cook" controls; it uses the single-chip microcomputer made by Texas Instruments.

For companies large and small, instrumentation is proving to be one of the most rewarding areas of microprocessor applications. Because of its powerful data-processing capacity, a computer-on-a-chip can not only impart brand-new capabilities to an instrument but also make it much easier to operate. With the microcomputer helping out, an unskilled person can operate a complex instrument, because, as one Perkin-Elmer engineer puts it: "The skill now resides in the microcomputer." Perkin-Elmer has already introduced two different spectrophotometers incorporating the microcomputer and is working on other uses in scientific instruments.

Microcomputers will also make a lot of laboratory-type analytical equipment more readily applicable to process control. Leeds & Northrup has already produced one such instrument, a particle analyser that uses a laser beam to measure particles and a microcomputer to figure out their size distribution. The device is being tested in a taconite (iron ore) plant, but it can be adapted to other customers' needs through a change in its program.

Semiconductor manufacturers are also looking for applications of microcomputers to appliances such as washing machines and refrigerators. The current recession has delayed new-product introduction in this field, but microcomputers are being designed into models that are expected to start showing up in about two years.

The automobile may prove to be a big user of electronics in years to come. Some electronic components are already being employed in cars to supervise ignition, measure voltages, and so on. Microcomputers are expected to start appearing in automobiles towards the end of this decade. Ford Motor Co. has found that microcomputer-run controls can cut fuel consumption by as much as twenty per cent under test conditions. The company plans to introduce the tiny computers in a 1979 car. Other auto-makers have similar plans.

In many other areas, microcomputers promise spectacular advances. In the home, microcomputer controls could result in savings on electric and heating bills. For the military, the tiny computers promise the evolution of more versatile weapons. In medical electronics, they open up possibilities for compact and less costly diagnostic instruments. There are indications that in conjunction with complex optical and mechanical devices, microcomputers could help restore vision for some of the blind. In one project, a microprocessor chip will be embedded in an eyeglass frame to decode visual information from artificial "eyes" and send it to the brain.

As is true with any other computer, the largest costs – and most problems – arise in writing application programs for microcomputers. Basically, a digital computer runs in response to instructions written in the binary code of ones and zeros. That's how the first computers were programmed – with the complex instructions written out painstakingly by hand. To ease the programmers' task, the industry has over the years developed high-level computer languages in which abbreviations or even words substitute for whole series of numbers. Along with the languages came such programming aids as assemblers and compilers.

The semiconductor industry makes such aids available to microcomputer users. The machines are, in effect, small computers that utilize microprocessors. They sell for $2500 to $10,000. Motorola calls its device the Exorciser; Intel's is called the Intellec.

Problems arise when design engineers who have previously dealt with electromechanical relays, or even hardwired logic, and are untutored in computer programming, suddenly face the complex accoutrements of data processing. For some, says one specialist, the experience is like "going from wood burning to nuclear fuel". As a result, something of an occupational obsolescence has temporarily developed in the design field because the engineers who are most skilled in product design usually have little or no experience with microprocessors and their applications.

Trying to fill the educational gap, MIT and some other universities have begun intensive courses for both students and industry representatives. Reports MIT Professor H. M. D. Toong: "Students go right from here out into industry and get jobs first thing heading microcomputer development and applications departments." Some specialists think that the applications of microcomputers will start expanding manyfold when the new graduates begin to enter the work force in large numbers.

For semiconductor companies, the microcomputer opens another broad avenue for growth. With phenomenal price declines a way of life, the industry is a voracious consumer of new markets. Industry executives like to note that the price of an electronic function such as a transistor dropped 99.9 per cent from 1960 to 1970 and is still declining. As one man puts it: "It's like putting an $8 price tag on an $8000 Cadillac."

At the same time, each new advance in technology has brought with it a widening use of electronics. Texas Instruments calculates that during the vacuum-tube era, digital-electronic sales rose on a slope of about ten per cent a year. In the days of the transistor, the slope steepened to an eighteen per cent annual increase. Integrated circuits increased the sales growth rate to thirty-eight per cent. Now TI expects another upward tilt in the curve in the late 1970s, thanks chiefly to the microcomputer. The company anticipates that for the foreseeable future sales of electronic components will climb at a dizzying rate of fifty to sixty per cent a year.

There seems to be little disagreement that the microcomputer is close to being an ultimate semiconductor circuit and that it now sets the direction for semiconductor technology. On the face of it, the principal beneficiary of this trend would appear to be Intel. The company now dominates the microcomputer market. What's more, it mainly makes semiconductor memories of the kind that go into microcomputers and does not make the integrated circuits that microcomputers replace. The principal losers would seem to be Texas Instruments, Fairchild, Motorola, and National Semiconductor, which are big in what is called transistor-transistor logic (TTL), the mainstay of the integrated-circuit business today – precisely the circuits the microcomputer replaces.

But that's not how top executives of some of those companies see the future. TI's Fred Bucy envisages his company emerging as a major force in microcomputers. So does Charles E. Sporck, president of National Semiconductor. And both are probably right. Bucy stresses, and others agree, that the microcomputer's biggest use will be in applications where electronic devices have never been employed before. New applications thus will be far more important than replacement of TTL logic. Bucy also notes that TI is the only semiconductor company "that has lived through all the generations of electronic components. We've successfully moved from one horse to the next." Few executives in the industry would dispute TI's obvious strengths as a $1.5-billion company even if it has been late in microcomputers. National Semiconductor, too, is an exceedingly clever marketer.

Everyone agrees, furthermore, that there will be a whole spectrum of microcomputers aimed at different applications, with many companies sharing the anticipated big market. And it is generally agreed that the most successful makers of microcomputers will be those that supply the best operating programs. The need to generate software to go with the tiny computers is a new activity for semiconductor companies, with the exception of TI, which for years now has been making both minicomputers and very large machines.

Bucy and other TI executives feel that's another plus for their company. To keep its computers tied together, and to ease the task of users who want to employ microcomputers in conjunction with bigger machines, TI early in 1975 introduced a powerful microcomputer whose software is compatible with that of the company's minis. TI sees a big competitive advantage in this approach, since the software of most other microcomputers does not directly match that of bigger computers.

The ability of users to operate a whole hierarchy of computers, from a big host machine to the microcomputer far down in the organization, will speed the trend towards "distributed" computer power. Bucy sees as a result a computer world polarized into giant machines and huge numbers of microcomputers, with medium-sized computers diminishing in importance.

Other specialists see computers of the future evolving into modular processor systems based on microcomputers, with many of their programs embedded in microcomputer memories, replacing expensive software. Frederick G. Withington, of Arthur D. Little, Inc., predicts that ten years from now, as a result of the semiconductor industry's nonstop price erosion, the cost of even the largest CPU may come down to about $30,000.

Manufacturers of bigger mainframes are indeed beginning to incorporate microcomputers not only into terminals and minicomputers but also into their large machines, to control such functions as input and output of data. A vice-president of NCR says that his company is "going to concentrate on the use of microprocessors in microcomputers, minis, and on up the line". NCR buys microcomputers from semiconductor manufacturers but it also plans to make its own. Burroughs already manufactures its own microcomputers and uses them in a variety of devices, including a small business computer. Control Data buys from Intel. IBM and Honeywell do not yet make a microprocessor on a chip.

For manufacturers of big mainframes, then, the microcomputer has so far been a new component rather than a competitor. But for manufacturers of minicomputers, the arrival of the microcomputer has created a competitive danger – the micros are encroaching on the minis. To counter the threat, Digital Equipment Corp., No. 1 in minis, has made arrangements with a semiconductor company, Western Digital Corp., under which Western makes microprocessors and associated components. Digital Equipment then puts the devices on circuit boards and sells the microcomputers in direct competition with the semiconductor houses.

Some semiconductor companies, in turn, have come out with microcomputers that run on programs written by Digital Equipment and Data General Corp. for their minicomputers. These microcomputers do essentially the same job, but sell for a lot less than the original minis. This blurring of dividing lines between computer and semiconductor manufacturers is expected to continue. Only half in jest, Noyce already calls Intel "the world's largest computer manufacturer".

In its impact, the microcomputer promises to rival its illustrious predecessors, the vacuum tube, the transistor, and the integrated-circuit logic chip. So far, probably no more than ten per cent of the tiny computer's potential applications have reached production stage. Today, nearly thirty years after the debut of the ENIAC, there are about 200,000 digital computers in the world. Ten years from now, thanks to the microcomputer, there may be twenty million.

The Electronics Revolution

Philip H. Abelson and Allen L. Hammond

Written by the editor and a staff member of Science *magazine, this article sets the microelectronic revolution in historical perspective and surveys the areas of social life it will affect. It formed the introduction to a special issue of* Science *magazine on 18 March 1977 (vol. 195, No. 4283).*

Earlier in the century, the United States experienced a long era of sustained growth in many aspects. There was a steady increase in level of education, life expectancy, and standard of living. Growth of all kinds was welcomed, including industrial expansion and population increase. A feeling of progress, of achievement, of well-being was everywhere. As a corollary, a striving for excellence and the search for understanding were widely admired.

Today the mood of America has turned pessimistic and negative. Those who are so inclined can find much evidence to support these views. Growth in the use of energy in the form of oil was suddenly curtailed in 1974. Growth in consumption of the kind seen in the 1950s and 1960s will not occur again. That part of the standard of living which is based on large-scale consumption of energy is not likely to improve during this century.

However, those who prefer optimism have reason for hope. Human ingenuity in solving problems is great. And native intelligence has been amplified enormously by the use of knowledge accumulated through research. An important product of research and a basis for hoping for a bright new future is the vitality of the electronics revolution. This revolution has been in progress for about sixty years. Lately its tempo has increased greatly. Until recently, its importance was overshadowed by changes due to the large-scale expansion in the use of energy. But it promises to be more important, of more enduring consequence, than the earlier industrial revolution.

Some of the great changes brought about by the electronics revolution have gone comparatively unnoticed; at the start they were evolutionary rather than sudden and drastic. The telephone, which we take for granted, was invented one hundred years ago. Nearly every decade since then the quality and scope of service have steadily improved, and the cost (measured in constant dollars) is now a tiny fraction of what it was fifty years ago.

Numerous applications of electronics gradually affected individuals and almost every component and activity of society. Radio was a marvellous toy and a source of wonderment when it was introduced fifty years ago. Now Americans listen to commercial radio an average of nearly four hours daily, and radio is accepted as practically a natural phenomenon. Television, which created a stir twenty-five years ago, is likewise commonplace.

During the last few years the impact of electronics on society has increased greatly. Examples are the rapid growth in popularity of citizens band radio, the worldwide use of the telephone, and the current astonishingly low prices at which handheld calculators and electronic watches are being sold. Less evident to the individual but in total more important to society are other applications of electronics that affect nearly every sector of our economy.

This revolution, which is destined to have great long-term consequences, is quite different in nature from the industrial revolution. The industrial revolution was based on a profligate use of energy (mainly fossil fuels). Much of its technology was crude, with only a modest scientific or theoretical base. In large measure what the industrial revolution did was to make available and to employ large amounts of mechanical energy.

In contrast, the electronics revolution represents one of the greatest intellectual achievements of mankind. Its development has been the product of the most advanced science, technology, and management. In many applications, electronics requires little energy. Indeed, one of the factors that guarantee enduring impact for the electronics revolution is that it is sparing of energy and materials.

With electronics one can control the disposition of large amounts of energy and force, but much in the way the brain is used in directing the action of muscles. In some aspects, electronics can be more subtle, more nimble, more dependable than the brain. In other applications, electronics serves as a great extender of human capabilities by rapidly carrying out routine but complex calculations, thus freeing the mind to make intuitive judgements and find shortcuts to new insights.

The industrial revolution, dependent on energy and materials, will be slowed and limited by the paucity of these necessary ingredients. The electronics revolution, fuelled by intellectual achievements, is destined for long-continued growth as its knowledge base inevitably increases. Obviously, the current rapid rate of evolution of electronics cannot persist indefinitely, but significant change is likely to continue for a long time.

One of the factors contributing to this dynamism is that in laboratories devoted to extending the electronics revolution the use of powerful investigative tools based on electronics is speeding new developments. Moreover, the body of knowledge that is being accumulated in the natural sciences continues to grow, and its growth has been fostered by the new tools that electronics has provided. There are few laboratories devoted to studies on the frontiers of the natural sciences that are not dependent on one or many items of electronic-based equipment. Two examples indicate the extent of the impact. A human's speed of reaction is about one fifth of a second. Measurements can now be made in times as short as 10^{-12} second. More important is the overall effect of electronic devices on quantitative determinations of many kinds. In some instances, sensitivities have been increased by orders of magnitude while the times required for measurement have been diminished to a hundredth or less of those needed in earlier methods.

One of the factors favouring the development of electronics has been a comparatively high degree of social acceptance. There have been sporadic attacks on various electronic devices such as computers and there is continuing concern about privacy, but the intensity of criticism has diminished. In comparison to the number of objections raised to chemical products, to the environmental concerns associated with nuclear and fossil fuel energy, or to fears of recombinant DNA, objections to electronics have been few.

The average citizen is fearful of air pollution, for example, and is frustrated by a feeling that there is little an individual can do about it. In contrast, if a television programme is offensive, it can be summarily dispensed with. Items that have recently become broadly available, such as the handheld computer, electronic watch, and citizens band radio, enhance the public's feeling of participating in the benefits of electronics while not bringing with them discernible side effects. In future, electronics will provide many new tools useful to the general public.

Three major themes are worth stressing: the tempo of the revolution, its magnitude, and the changing driving forces that have spurred it.

Until 1940, developments in electronics took place at a comparatively moderate pace. As was true with many scientific and technological matters, the pace quickened during World War Two and was further maintained during the Cold War. Two major developments occurred independently during the late 1940s and later fused to give enormous impetus to electronics. One was the construction of programmable electronic computers. The second was the invention of the transistor. Subsequent developments in solid state physics led to the present-day silicon chip with its large-scale integrated circuits. One such circuit can contain more active elements than the most complex equipment of twenty-five years ago.

After about 1960, when solid-state devices were incorporated in computers, there was a rapid development in the capabilities of computers and a steady reduction in the costs of calculations. An important effect of integrated circuits has been a reduction in the size and power requirements of electronic equipment that has made possible, for example, a Viking lander. Other advantages include reproducibility, maintainability, and reliability. Especially helpful is a sharp decrease in the need for making interconnections.

The tempo of change has been impressive. In 1959, a chip that was commercially available contained one component of a circuit. By 1964, the number of components per chip had risen to ten, by 1970 to about 1000, and by 1976 to about 32,000. The cost per chip advanced only modestly. Thus, the cost per function has dropped drastically. It is this great change in the cost-effectiveness ratio that has made possible inexpensive handheld calculators and related microprocessors and minicomputers. One of the key individuals who have been pushing the development of large-scale integrated circuits is Robert Noyce, the co-founder of Intel Corp. He argues that further advances can be expected. Theoretical considerations show that physical limits have not been approached. He is so bold as to state that "if the present rate of increase of complexity were to continue, integrated circuits with 10^9 elements would be available in twenty years."

All of us have seen examples of the sudden termination of exponential growth, so perhaps Noyce's figure will never be attained. But substantial advances toward his goal are already in progress. He seems justified in the view that "the potential for developing inexpensive processing power is truly awesome." He projects that with low-cost processing many new tasks will be undertaken that are uneconomical today.

Another way of glimpsing the tempo and magnitude of the electronics revolution is to focus on what has been happening in computers. In the early 1950s, almost all computers were owned by or devoted to tasks of the federal government. Computers were procured for use in such applications as defence and nuclear reactor design. By the mid-1950s there were about 1000 large-scale computers, and the tendency was toward increasing computational power. By the mid-1960s, there were 30,000 computers, and the generally accepted view was that costs of computation decreased with size; that is, the larger the better. At the end of 1976 there were about 220,000 computers in the United States. Of these, forty per cent were medium or large computers; the remainder were minicomputers which are small and by definition cost less than $50,000. At the same time, there were 750,000 of the microprocessors that form the heart of microcomputers. Ruth Davis has estimated that by 1980 the number of minicomputers will reach 750,000, while the number of microprocessors will increase to more than 10 million.

As the number of computers in service grew, the uses and the organizations involved broadened. The current distribution of ownership of conventional computers, with the percentages in the major categories, are: manufacturing industry, 31; miscellaneous business, 13.3; financial institutions, 13.4; wholesale and retail trade, 13.1; educational institutions, 5.7; state and local government, 5.7; and federal government, 3.4. There is further scattering of ownership throughout virtually every kind of organized activity. Thus it may seem that an enormous shift in the nature of the market for large computers has occurred. Beyond that is the larger market for minicomputers and the much larger mass demand for microprocessors.

Coincident with the expansion in the number of computers has been an increase in the number of computer professionals. During the past twenty years the total number of analysts, designers, programmers and operators has increased from 100,000 to 2.5 million. The number of students having some degree of familiarity with computers is much greater. This reservoir of people familiar with applications of computers is certain to facilitate additional applications of electronics. The emergence of computer hobby shops is bringing additional enthusiasts and imagination into the field. One group that is likely to make substantial contributions is the working scientists in the natural sciences. Often their progress and ability to tackle problems are limited by their equipment. Having experienced the advantages of incorporation of microprocessors in measuring devices, they will be looking for novel kinds of electronic sensors that can be coupled with the current data processors.

Because many of the new major applications involve various kinds of computers, one might have the impression that the electronics revolution and computers are synonymous. It is easy to lose sight of the importance of the noncomputer aspects of electronics. Key to many applications are the transducers or sensors. For example, computers would have a limited role in process control if electronic devices for sensing temperature, pressure, and concentrations of components were not available.

The potential applications of electronic technologies are so numerous and so provocative as to give free rein to futurologists and science-fiction writers. The domestic robot, the wired city, the global electronic village – none of these can be dismissed as being beyond the bounds of technical feasibility. But it is not necessary to look so far afield to see how pervasive the impact of electronics is, how many areas of human endeavour and how large a portion of the country's economic activity may be substantially altered. Indeed, it is probable that reality will outstrip fiction in the rate of introduction of new and often unexpected applications of electronics in coming years. Witness, for example, the incredible growth in popularity of citizens band radio. It is clear that the capability of some electronic devices, particularly microprocessor circuits and memory units on single silicon chips, is developing more rapidly than applications can be conceived of and introduced.

The markets for kitchen appliances, office equipment and leisure games, to mention just a few, are ready to be revolutionized or at least substantially modified by the addition of logic and memory to yield "smart stoves" and similar products, the first of which are already available.

The driving force behind many of the commercial applications is the extremely low prices for sophisticated electronic circuits, which in turn derive from mass production. The key innovation allowing such large markets is the microprocessor, a general-purpose logical unit that can be programmed to perform an unlimited number of tasks, thus eliminating the necessity of designing new circuitry for each new application. Among other applications, microprocessors are making it possible to extend computer control to mechanical and electrical equipment of every description, from consumer appliances and automobile engines to milling machines and industrial boilers. In the past, automation of manufacturing and process control has moved slowly because of fear of dependence on a central computer and the cost of the controlling units. The first process control computers introduced in the late 1950s, for example, cost about $300,000; minicomputers reduced this to less than $100,000 by the late 1960s; now microprocessor controllers are available for $3000, cheap enough to automate control and data collection for even small process steps. What seems to be evolving is a linked, hierarchical arrangement in which microprocessors are used to control individual pieces of equipment; minicomputers collect and process management information from the microprocessors for an entire factory; and large central computers use the resulting data in compiling corporate financial reports.

But the impact will not be confined merely to consumer products or isolated devices. The application of electronics is already having a pervasive effect on the entire economy and on our way of life, one that promises to intensify in coming years. Consider the following areas – medicine, education, national defence, banking and retail sales, postal and other communications, and the research process itself.

The practice of medicine, for example, has already begun to change in such areas as the handling of patient records, billing and other administrative chores, and computer-controlled examinations in response to conventional data-processing equipment. Even more fundamental extensions of the physician's skill are resulting from the application of compact integrated circuitry to diagnostic and monitoring equipment. The potential of medical electronics is indicated by the unprecedented demand for tomographic X-ray scanning equipment, which by computer processing and synthesis is able to distinguish different tissues with a sensitivity fifty times that of ordinary X-ray techniques; hundreds of these new diagnostic tools have been ordered. Another new and non-invasive diagnostic approach, the use of acoustic waves in such devices as ultrasound cameras, is also beginning to be widely applied; here the key role of electronics is to translate the acoustic information into visual and analytical data. Perhaps the most striking illustration of the unique power of electronic circuits in medicine is their potential use as prostheses to supplement or replace damaged neural tissue, a circumstance that is only possible because modern circuits now approach the size, power consumption, and logical capability of the natural tissue. The cardiac pacemaker is an early example of such a prosthesis and the development of far more complicated devices such as an implantable electronic ear for the deaf is well under way.

Potentially, electronics and electronic media could have an important impact on education, as almost anyone who has observed children watching Sesame Street could confirm. Despite a few such successes, however, there seems to be general agreement that television and computer-assisted instruction have not yet lived up to that potential. But educational innovators have not yet given up. There is another sense, however, in which electronics is certain to affect education for better or for worse, and that concerns the prospective flood of inexpensive electronic devices of which the handheld calculator is only the first. Calculators have substantially altered the character of the traditional "problem sets" in science and engineering courses at the university level, they are becoming common in high school courses, and they are already creeping into use in primary schools. Some parents and educators are trying to stem this growing tide on the grounds that it will only add to the reasons "why Johnny can't add". Others see the trend as inevitable and point out that how computation is performed is irrelevant, what really counts is whether the students learn the underlying concepts, and in this respect the impact of the calculator is still uncertain. In any case, the ubiquity of the calculator seems to guarantee that electronic arithmetic will become the language of the real "new maths", and these developments suggest the potential of the more elaborate calculator-based games and educational devices that are beginning to appear either alone or as attachments to the home television set.

It is difficult to imagine a modern military force without heavy dependence on electronics. Aircraft instrumentation, missile guidance systems, radar and other surveillance sensors, tactical computers – all depend on electronic components. But electronics plays more than a passive role in military systems; in recent years, advances in electronics have been perhaps the most important factor guiding the evolution of new weapons and new strategies. One example is the emergence of "smart bombs" and other unpiloted weapons, which can evaluate guidance information to track themselves to target or make use of pre-programmed instructions to manoeuvre evasively. Carried to its logical conclusion, this trend might eliminate the need for many manned aircraft and is at the core of the current debates over the B-1 bomber and the cruise missile. A second example is the NAVSTAR satellite system, for which prototypes are now being tested. These navigational satellites are designed to allow any military vehicle carrying an inexpensive receiver and computer to instantaneously determine its position anywhere in the world with an accuracy of better than ten metres in horizontal and vertical coordinates; civilian aircraft and ships will also be able to use these satellites, but with somewhat less accuracy. This phenomenal accuracy is expected by many defence analysts to revolutionize navigation, weapons targeting, battlefield management, and other aspects of warfare; this is especially true for fixed targets, since the coordinates of any such target can be readily determined. The system also may supplant many of the commercial navigation systems now in use.

In twenty years, the role of computers in research has been transformed from what W. O. Baker has described as "a minor annex of mathematics research" to a major and often dominant role characterized by the proliferation of minicomputers and timeshared terminals in most research institutions. He asserts that computers have transformed the research process from conceptualization to experimentation to publication. Baker should know, since his organization has been in the forefront of actually putting computers to work in research – Bell Laboratories now have an average of one dedicated minicomputer and five interactive terminals for every fifteen professional staff members. Electronics has also transformed other instruments of scientific research, from the electron microscope to vidicon astronomical cameras, and the process is accelerating as more and more "smart instruments" are designed around microprocessors.

Nowhere is the potential impact of electronics greater than in the banking industry and the postal service, both of which face the prospect of converting from moving pieces of paper around to using electronic transfers for at least part of their business. These changes will certainly not come overnight and will raise a host of social problems; how does one protect against theft when most retail transactions are done electronically rather than by cheque, or guarantee privacy if much first-class mail travels by wire? But electronic transfers of money, of messages, and of documents are already established features of our society. The US government, for example, makes some five million social security payments each month by sending banks magnetic tapes that the bank computers are able to use to credit depositors' accounts directly. Pre-authorized, nonpaper payments are estimated to account for ten per cent of bank transactions in some areas of the country. Point-of-sale electronic terminals are now becoming common in retail stores, although their role is presently restricted to credit verification and inventory control, not direct transfer of funds from the customer's account to the store's. But with twenty-six billion cheques a year passing through the banking system and the likelihood that the volume will double by 1985, there is ample incentive for banks to move towards electronic transfer and chequeless banking. Such a move will change the boundaries between the retail trade and the banking system, possibly resulting in more decentralized but far more complex financial networks.

Chequeless banking, if and when it does occur, will intensify the economic pressures on the postal service, since nearly forty per cent of the mail consists of cheques and other financial transactions. Diversion of this mail will reduce revenue but will not noticeably lower costs. But that is not the only threat. As complaints about lost or delayed mail increase, many large businesses are looking towards electronic mail systems. One prototype of such a system is the Department of Defence's ARPA computer network, which is routinely used by researchers all across the country to exchange messages and information. New optical scanning and electronic printing techniques are being developed by many companies that would allow users to transmit documents or whole pages of text. The postal service may be forced to embrace electronic mail or face the future as an obsolete, increasingly expensive system serving fewer and fewer people.

Postal communications are not the only form of communications facing new challenges. Not long ago radio and marine cable telephone circuits were the principal means for rapid intercontinental communications. Now an international satellite communications system is well established and carrying a growing volume of traffic. Domestic satellite systems are just getting under way in the United States, but they seem certain to expand the options for voice, television, and digital data communications. The transmission of digital information is the most rapidly growing area of electronic communications and reflects the increasing need for computers and other intelligent machines to "talk" with each other. Indeed computer and communications technologies have become so similar and intertwined that they are difficult to distinguish. More and more communications are transmitted in digital form, even within the telephone system. And more and more information in the communications systems is processed both before and after transmission; voice signals are compacted and compressed to put more calls on a channel, for example, or the output of the intelligent terminal on one end of the phone line becomes the input for a computer or a display device on the other end. Distributed processing – essentially networks of small and medium-sized computers connected by communications links – is clearly going to be one of the major forms in which computers are used.

These developments clearly pose a major problem for those, such as government regulators, who must decide where communications – a regulated activity – ends and where the unregulated computer market begins. There are a number of public policy issues involved in this intensifying conflict, and among the principal contenders are some of the giants of American industry, AT&T and IBM. It promises to be a multibillion-dollar fight and one of the thorniest technological policy problems the government must face in the coming decade.

The evolution of computers to the point where communications is a major part of their activity is also reflected in other changes. The traditional use of computers as calculating engines for numerical work is rapidly being replaced by a new principal role, that of managing, storing, retrieving, and distributing information. A search for new computer architectures that better reflect this new role is under way and includes experiments with such things as augmented sets of instructions for the computer's own control program, specialized subcomputers or processors to manage data-bases, and multiprocessor machines. The generation of computer programs – all too often a bottleneck to effective use of computers – is increasingly being put on a firm mathematical basis. The way information is stored in computer systems is changing too, as researchers look for more efficient search routines, new methods of combining memory devices, and new computer languages adapted for information processing. S. E. Madnick believes that what is emerging is the goal of an information utility that can serve many users for many purposes. Prototype experiments with information utilities are actually under way in Britain, where the British Broadcasting Corporation is testing a small decoder attached to television sets that can, on demand, deliver current information on a variety of subjects; the British Post Office, which runs the telephone network, is also experimenting with a telephone-based information utility that would combine the functions of a daily newspaper with the resources of a library.

We are increasingly an information-based society. Economically, for example, information industries ranging from broadcast television to book publishing to computer services contribute a large part of the US gross national product and employ nearly half of the work force. The information sector of the economy is also among the more rapidly growing. Moreover, information is a resource that greatly enhances individual capabilities and opportunities and is not depleted by use. H. A. Simon argues that the development of the ability to process and manipulate information on a large scale has a significance, on the scale of human evolution, equivalent to the development of written language or the invention of the printed book. In any case, it is clear that the information revolution will accelerate as more persons acquire their own computers and as these computers are able to make use of larger and larger information resources.

The conventional economic wisdom is that the expanding opportunities in information-related activities will more than offset jobs lost to more productive electronic equipment. Recent experience in the electronic manufacturing industries would seem to bear this out, although substantial layoffs in some companies have been avoided only by a commitment to large retraining schemes. In the coming decade automated electronic equipment is likely to make inroads into the service sectors of the economy as well. The nature of secretarial and other office support jobs may change, for example, as may that of mail clerks and bank tellers. Whether any of these changes will result in displacing large numbers of people from these traditionally labour-intensive occupations is not clear, but the process of change is certain to be uncomfortable for the individual whose job is involved and maybe for society as a whole. Consider the impact of converting the postal service, one of the largest employers of unskilled and semi-skilled labour, to an electronic mail system.

Despite such problems, the electronics revolution is not likely to slow down anytime soon, if only because the research base is broad and vigorous and is already producing a host of new ideas and new concepts that are certain to be translated into new products and services in coming years. One trend that can be identified is the incorporation of magnetic, acoustical and optical phenomena into electronic devices, giving rise to a host of new effects that can be put to use. Examples include magnetic bubbles in memory devices, surface acoustic-wave filters in signal processing equipment, and optical fibres in communications. Particularly fascinating in this regard is the push towards optical communications.

The concept of transmitting information on a light wave dates back at least to Alexander Graham Bell, who in the 1870s demonstrated a wireless telephone based on light that could transmit sound for more than a kilometre. A hundred years later, optical communication is on the verge of becoming a reality. The principal advantage of communicating with light waves is their high frequency, compared with radio waves, which gives them a superior capacity to transmit information. A single optical fibre, for example, can carry hundreds of times as many bits of information per second as a copper wire. The Bell system is already experimenting with an optical link for interstation connections in areas with a high volume of calls, and others are looking at applications ranging from computers to military vehicles. As large-scale production of fibres and other components gets under way, cost reductions as dramatic as those in electronic calculators are expected.

In its modern configuration, optical communications relies on such components as lasers or light-emitting diodes for light sources and glass fibres for the transmission medium. It is, in fact, the dramatic improvement in these components that has brought the technology to the verge of utility. Optical fibres have been produced with losses less than 1 decibel per kilometre – less than the loss in light passing through a single windowpane. But, just as solid-state technology progressed from individual devices to integrated circuits, integrated optical circuits incorporating lasers, amplifiers and detectors on a single chip are already being developed. Many semiconductor materials, it turns out, are optically as well as electrically active, thus permitting the intimate interplay of both kinds of circuits. This activity is leading to an era in which the electron, long the workhorse of the electronics revolution, will be supplemented by the even greater potential of the photon.

The range of phenomena and the indications of still-to-be-exploited potential to be found in electronics research are convincing evidence that we have not yet seen the limits of what is possible. Still less are most of us and most of our institutions prepared to decide what we should do with our new capabilities or even how to cope with the speed at which electronics technology is changing the ground rules under which we operate. In business the price of being unprepared is often high, as many white goods (cookers, fridges, freezers, washing machines, etc) manufacturers found out when one of their competitors introduced an electronically controlled microwave oven that has rapidly become the best-selling product in its field. For governments and individuals alike the stakes are arguably lower at present, but the continuing electronics revolution promises to be so pervasive as to compel the attention of even the most unobservant.

Microelectronics

Robert N. Noyce

Perhaps the most famous of all statements on microelectronics, by a founder of the most famous of all microprocessor manufacturers, Intel Corp., this paper appeared in a special issue of Scientific American *in September 1977 (vol. 237, No. 3). Noyce's graphs illustrating the basic "laws" of microelectronics have been widely copied and quoted, testifying to the fact that this is indeed a definitive article.*

The evolution of electronic technology over the past decade has been so rapid that it is sometimes called a revolution. Is this large claim justified? I believe the answer is yes. It is true that what we have seen has been to some extent a steady quantitative evolution: smaller and smaller electronic components performing increasingly complex electronic functions at ever higher speeds and at ever lower cost. And yet there has also been a true revolution: a qualitative change in technology, the integrated microelectronic circuit, has given rise to a qualitative change in human capabilities.

It is not an exaggeration to say that most of the technological achievements of the past decade have depended on microelectronics. Small and reliable sensing and control devices are the essential elements in the complex systems that have landed men on the moon and explored Mars, not to speak of their similar role in the intercontinental weapons that dominate world politics. Microelectronic devices are also the essence of new products ranging from communications satellites to handheld calculators and digital watches. Somewhat subtler, but perhaps eventually more significant, is the effect of microelectronics on the computer. The capacity of the computer for storing, processing and displaying information has been greatly enhanced. Moreover, for many purposes the computer is being dispersed to the sites where it is operated or where its output is applied: to the "smart" typewriter or instrument or industrial control device.

The microelectronics revolution is far from having run its course. We are still learning how to exploit the potential of the integrated circuit by developing new theories and designing new circuits whose performance may yet be improved by another order of magnitude. And we are only slowly perceiving the intellectual and social implications of the personal computer, which will give the individual access to vast stores of information and the ability to learn from it, add to it and communicate with others concerning it.

Here I want primarily to show how the evolution of microelectronics illustrates the constant interaction of technology and economics. The small size of microelectronic devices has been important in many applications, but the major impact of this new technology has been to make electronic functions more reproducible, more reliable and much less expensive. With each technical development costs have decreased, and the ever lower costs have promoted a widening range of applications; the quest for technical advances has been required by economic competition and compensated by economic reward.

It all began with the development thirty years ago of the transistor: a small, low-power amplifier that replaced the large, power-hungry vacuum tube. The advent almost simultaneously of the stored-program digital computer provided a large potential market for the transistor. The synergy between a new component and a new application generated an explosive growth of both. The computer was the ideal market for the transistor and for the solid-state integrated circuits the transistor spawned, a much larger market than could have been provided by the traditional applications of electronics in communications. The reason is that digital systems require very large numbers of active circuits compared with systems having analog amplification, such as radios. In digital electronics a given element is either on or off, depending on the input. Even when a large number of elements are connected, their output will still be simply on or off; the gain of the individual stage is unity, so that even cascading several stages leaves the gain still unity. Analog circuits, on the other hand, typically require amplification of the input. Since the gain of each amplifier may typically be ten, only a few stages can be cascaded before the practical limit of voltage levels for microelectronic elements is reached. An analog system therefore cannot handle large numbers of microcircuits, whereas a digital system requires them; a pocket calculator contains one hundred times as many transistors as a radio or a television receiver.

In spite of the inherent compatibility of microelectronics and the computer, the historical fact is that early efforts to miniaturize electronic components were not motivated by computer engineers. Indeed, the tremendous potential of the digital computer was not quickly appreciated; even the developers of the first computer felt that four computers, more or less, would satisfy the world's computation needs! Various missile and satellite programs, however, called for complex electronic systems to be installed in equipment in which size, weight and power requirements were severely constrained, and so the effort to miniaturize was promoted by military and space agencies.

The initial approach was an attempt to miniaturize conventional components. One program was "Project Tinkertoy" of the National Bureau of Standards, whose object was to package the various electronic components in a standard shape: a rectangular form that could be closely packed rather than the traditional cylindrical form. Another approach was "molecular engineering". The example of the transistor as a substitute for the vacuum tube suggested that similar substitutes could be devised: that new materials could be discovered or developed that would, by their solid-state nature, allow electronic functions other than amplification to be performed within a monolithic solid. These attempts were largely unsuccessful, but they publicised the demand for miniaturisation and the potential rewards for the successful development of some form of microelectronics. A large segment of the technical community was on the lookout for a solution of the problem because it was clear that a ready market awaited the successful inventor.

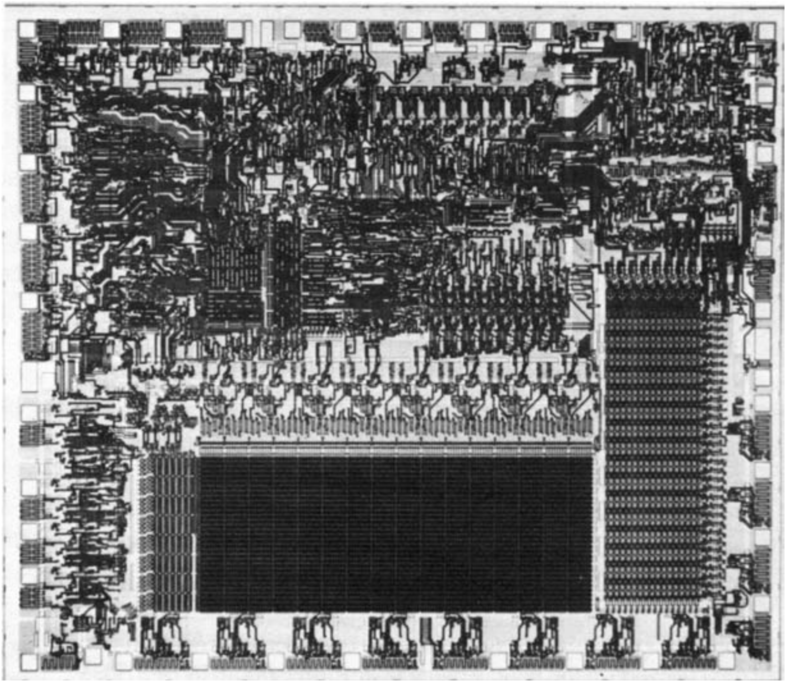
What ultimately provided the solution was the semiconductor integrated circuit, the concept of which had begun to take shape only a few years after the invention of the transistor. Several investigators saw that one might further exploit the characteristics of semiconductors such as germanium and silicon that had been exploited to make the transistor. The body resistance of the semiconductor itself and the capacitance of the junctions between the positive (*p*) and negative (*n*) regions that could be created in it could be combined with transistors in the same material to realize a complete circuit of resistors, capacitors and amplifiers. In 1953 Harwick Johnson of the Radio Corporation of America applied for a patent on a phase-shift oscillator fashioned in a single piece of germanium by such a technique. The concept was extended by G. W. A. Dummer of the Royal Radar Establishment in England, Jack S. Kilby of Texas Instruments Incorporated and Jay W. Lathrop of the Diamond Ordnance Fuze Laboratories.

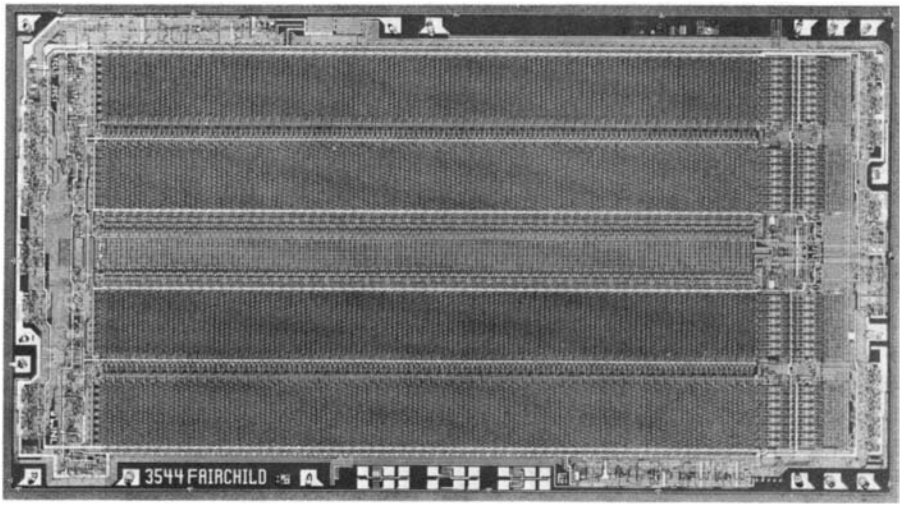
Several key developments were required, however, before the exciting potential of integrated circuits could be realised. In the mid-1950s engineers learned how to define the surface configuration of transistors by means of photolithography and developed the method of solid-state diffusion for introducing the impurities that create *p* and *n* regions. Batch processing of many transistors on a thin "wafer" sliced from a large crystal of germanium or silicon began to displace the earlier technique of processing individual transistors. The hundreds or thousands of precisely registered transistors that could be fabricated on a single wafer still had to be separated physically, assembled individually with tiny wires inside a protective housing and subsequently assembled into electronic circuits.

The integrated circuit, as we conceived and developed it at Fairchild Semiconductor in 1959, accomplishes the separation and interconnection of transistors and other circuit elements electrically rather than physically. The separation is accomplished by introducing *pn* diodes, or rectifiers, which allow current to flow in only one direction. The technique was patented by Kurt Lehovec at the Sprague Electric Company. The circuit elements are interconnected by a conducting film of evaporated metal that is photoengraved to leave the appropriate pattern of connections. An insulating layer is required to separate the underlying semiconductor from the metal film except where contact is desired. The process that accomplishes this insulation had been developed by Jean Hoerni at Fairchild in 1958, when he invented the planar transistor: a thin layer of silicon dioxide, one of the best insulators known, is formed on the surface of the wafer after the wafer has been processed and before the conducting metal is evaporated onto it.

Since then additional techniques have been devised that give the designer of integrated circuits more flexibility, but the basic methods were available by 1960, and the era of the integrated circuit was inaugurated. Progress since then has been astonishing, even to those of us who have been intimately engaged in the evolving technology. An individual integrated circuit on a chip perhaps a quarter of an inch square can now embrace more electronic elements than the most complex piece of electronic equipment that could be built in 1950. Today's microcomputer, at a cost of perhaps $300, has more computing capacity than the first large electronic computer, ENIAC. It is twenty times faster, has a larger memory, is thousands of times more reliable, consumes the power of a light bulb rather than that of a locomotive, occupies 1/30,000 the volume and costs 1/10,000 as much. It is available by mail order or at your local hobby shop.

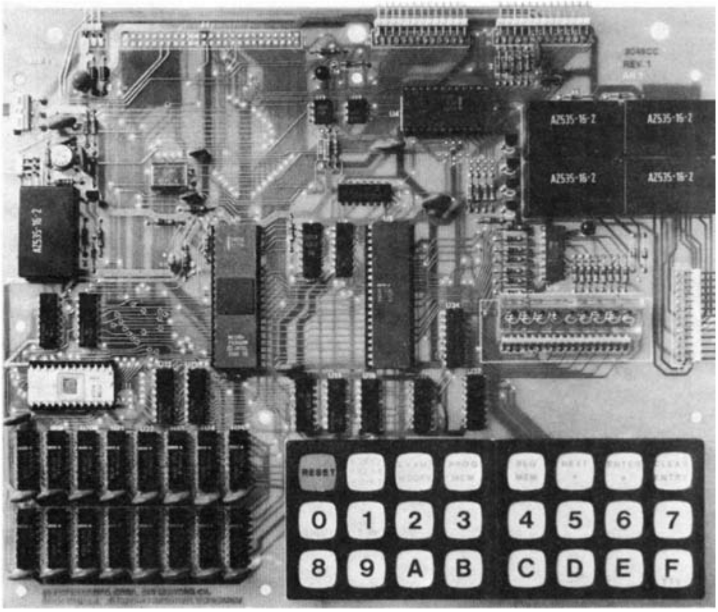
In 1964, noting that since the production of the planar transistor in 1959 the number of elements in advanced integrated circuits had been doubling every year, Gordon E. Moore, who was then director of research at Fairchild, was the first to predict the future progress of the





(Above) A blow-up of a large-scale integrated circuit made by Intel, which includes a memory and central processing unit. *Photo courtesy of Intel Corporation*

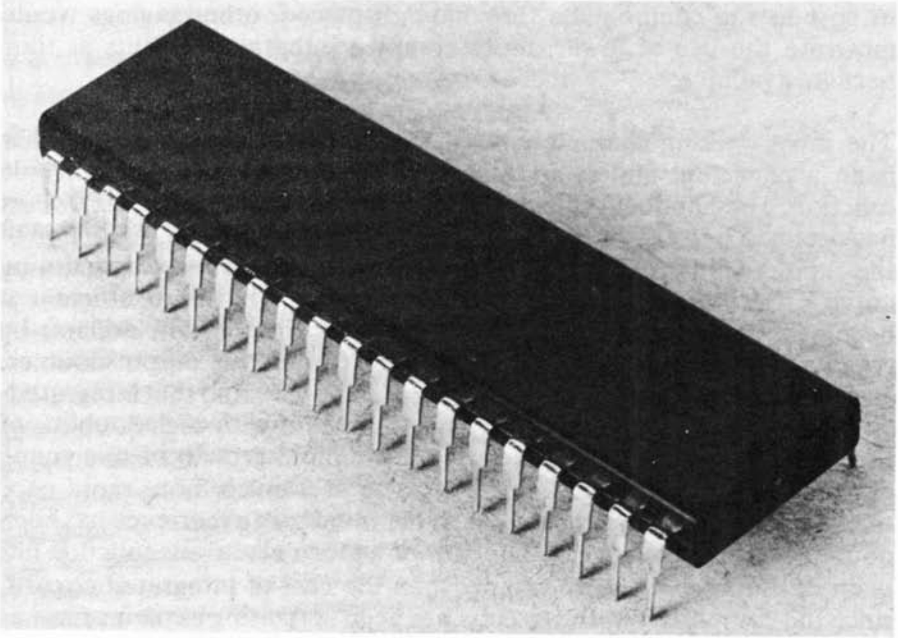
(Below) Another large-scale integrated circuit made by Fairchild, which consists entirely of memory. *Photo courtesy of Fairchild Camera and Instrument Corporation, 464 Ellis Street, Mountain View, California*



This IMSAI minicomputer measures 8 by 10 inches. The microprocessor is the square chip in the light grey package in the middle of the board array. *Photo by Ben Rose Photography, Inc., New York*

integrated circuit. He suggested that its complexity would continue to double every year. Today, with circuits containing 2^{18} (262,144) elements available, we have not yet seen any significant departure from Moore's law. Nor are there any signs that the process is slowing down, although a deviation from exponential growth is ultimately inevitable. The technology is still far from the fundamental limits imposed by the laws of physics: further miniaturisation is less likely to be limited by the laws of physics than by the laws of economics.

The growth of the microelectronics industry illustrates the extent to which investment in research can create entrepreneurial opportunity, jobs and a major export market for the US. After the introduction of the integrated circuit in the early 1960s the total world consumption of integrated circuits rose rapidly, reaching a value of nearly $1 billion in 1970. By 1976, world consumption had more than tripled, to $3·5 billion. Of this total, US-based companies produced more than $2·5 billion, or some seventy per cent, about $1 billion of which was exported to foreign customers. The impact on the electronics industry is far greater than is implied by these figures. In electronic equipment less



The familiar "caterpillar" profile of the packaged chip. *Photo courtesy of Ferranti Electronics Limited*

than ten per cent of the value is in the integrated circuits themselves: a $10,000 minicomputer contains less than $1000 worth of integrated circuits, and a $300 television set contains less than $30 worth. Today most of the world's $80 billion electronics industry depends in some way on integrated circuits.

The substitution of microelectronic devices for discrete components reduces costs not only because the devices themselves are cheaper but for a variety of other reasons. First, the integrated circuit contains many of the interconnections that were previously required, and that saves labour and materials. The interconnections of the integrated circuit are much more reliable than solder joints or connectors, which makes for savings in maintenance. Since integrated circuits are much smaller and consume much less power than the components they have displaced, they make savings possible in such support structures as cabinets and racks as well as in power transformers and cooling fans. Less intermediate testing is needed in the course of production because the correct functioning of the complex integrated circuits has already been ensured. Finally, the end user needs to provide less floor space, less operating power and less air conditioning for the equipment. All of this is by way of saying that even if integrated circuits were only equivalent in cost to the components they have displaced, other savings would motivate the use of fewer, more complex integrated circuits as they became available.

The most striking characteristic of the microelectronics industry has been a persistent and rapid decline in the cost of a given electronic function. The handheld calculator provides a dramatic example. Its cost has declined by a factor of one hundred in the past decade. A portion of the rapid decline in cost can be accounted for in terms of a "learning curve": the more experience an industry has, the more efficient it becomes. Most industries reduce their costs (in constant dollars) by twenty to thirty per cent each time their cumulative output doubles. Examining data for the semiconductor industry, we find that integrated-circuit costs have declined twenty-eight per cent with each doubling of the industry's experience. Because of the rapid growth of this young industry these cost reductions have come at a much more rapid pace than in mature industries; the electronics industry's experience has been doubling nearly every year. The cost of a given electronic function has been declining even more rapidly than the cost of integrated circuits, since the complexity of the circuits has been increasing as their price has decreased. For example, the cost per bit (binary digit) of random-access memory has declined an average of thirty-five per cent per year since 1970, when the major growth in the adoption of semiconductor memory elements got under way. These cost declines were accomplished not only by the traditional learning process but also by the integration of more bits into each integrated circuit: in 1970 a change was made from 256 bits to 1024 bits per circuit and now the number of bits is in the process of jumping from 4096 per circuit to 16,384.

The hundredfold decline in prices for electronic components since the development of the integrated circuit is unique because, although other industries have shown similar experience curves, the integrated-circuit industry has been unique in its annual doubling of output over an extended number of years. Rather than serving a market that grows only in pace with the gross national product or the population, the industry has served a proliferating market of ever-broadening applications. As each new application consumes more microelectronic devices more experience has been gained, leading to further cost reductions, which in turn have opened up even wider markets for the devices. In 1960, before any production of integrated circuits, about 500 million transistors were made. Assuming that each transistor represents one circuit function, which can be equated to a logic "gate" or to one bit of memory in an integrated circuit, annual usage has increased by 2000 times, or has doubled eleven times, in the past seventeen years. This stunning increase promotes continual cost reductions.

The primary means of cost reduction has been the development of increasingly complex circuits that lower the cost per function for both the circuit producer and the equipment manufacturer. The main technical barrier to achieving more functions per circuit is production yield. More complex circuits result in larger devices and a growing probability of defects, so that a higher percentage of the total number of devices must be scrapped. When the cost of scrapping exceeds the cost saving in subsequent assembly and test operations, the cost per function increases rather than decreases. The most cost-efficient design is a compromise between high assembly costs (which are incurred at low levels of integration) and high scrapping costs (which are incurred at high levels of integration).

Technological developments have concentrated primarily on increasing the production "yield", either by reducing the density of defects or by reducing dimensions. Meticulous attention to process control and

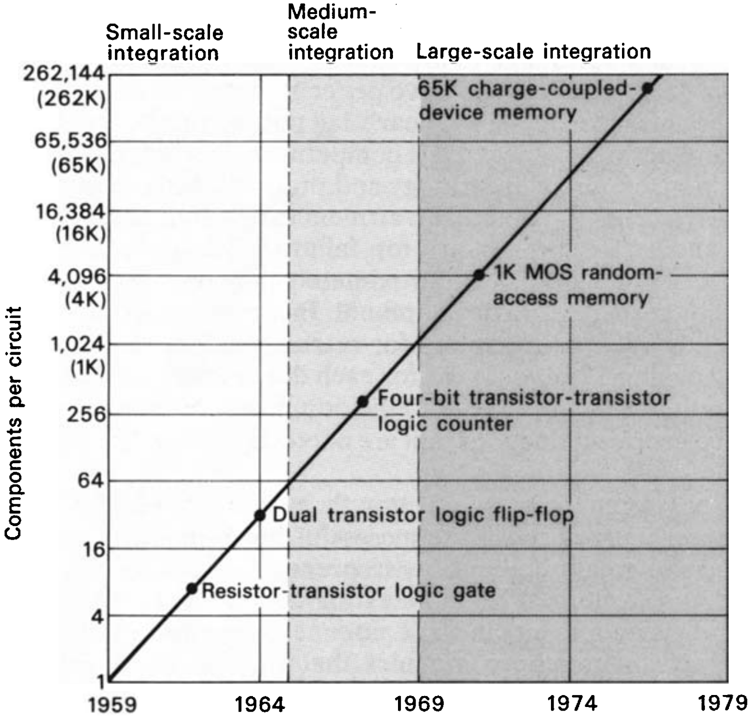


Fig. 1.1 Number of components per circuit

cleanliness has been necessary to reduce defect density. A dust particle in any critical process is enough to make a device worthless, so that most operations must be carried out in "clean rooms". Reduction of the dimensions of the basic circuit elements, which enables one to crowd more complex circuits within a given area, has been accomplished by improving the resolution of the photoengraving processes. Now optical limits are being reached as dimensions in the circuit patterns enter the range of only a few wavelengths of light, and methods in which electron beams or X-rays are substituted for visible light are being developed in order to reduce the dimensions even further (see William G. Oldham, below).

The reduction in size of the circuit elements not only reduces the cost but also improves the basic performance of the device. Delay times are directly proportional to the dimensions of circuit elements, so that the circuit becomes faster as it becomes smaller. Similarly, the power is reduced with the area of the circuits. The linear dimensions of the circuit elements can probably be reduced to about a fifth of the current size before any fundamental limits are encountered.

In an industry whose product declines in price by twenty-five per cent a year the motivation for doing research and development is clearly high. A year's advantage in introducing a new product or new process can give a company a twenty-five per cent cost advantage over competing companies: conversely, a year's lag puts a company at a significant disadvantage with respect to its competitors. Product development is a critical part of company strategy and product obsolescence is a fact of life. The return on successful investment in research and development is great, and so is the penalty for failure. The leading producers of integrated circuits spend approximately ten per cent of their sales income on research and development. In a constant-price environment one could say that investment for research and development buys an annuity paying $2·50 per year for each dollar invested! Clearly most of this annuity is either paid out to the purchasers of integrated circuits or reflected in price reductions that are necessary to develop new markets.

In this environment of rapid growth in market, rapid technological change and high returns on the successful development of a new product or process, a great number of entrepreneurial opportunities have been created and exploited. It is interesting that whereas the US has led in both the development and the commercialization of the new technology, it was not the companies that were in the forefront of the vacuum tube business that proceeded to develop its successor, the transistor. Of the ten leading US producers of vacuum tubes in 1955, only two are among today's top ten US semiconductor producers: four of

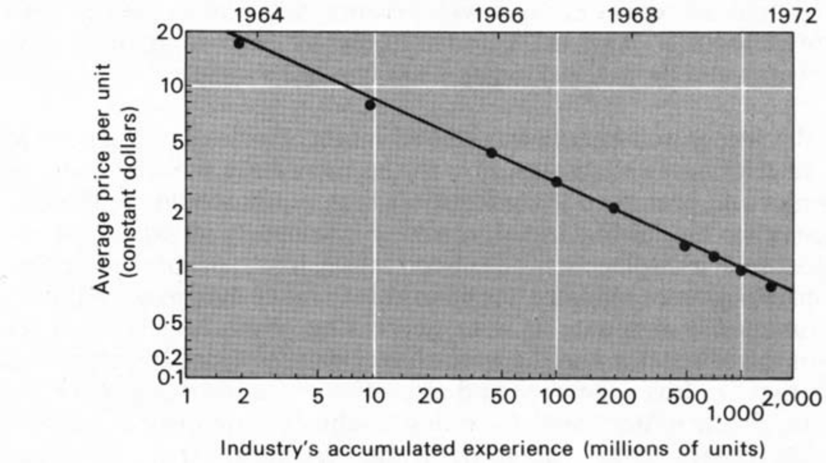


Fig. 1.2 Prices of integrated circuits

the top ten semiconductor companies were formed after 1955, and those four represent only a small fraction of the successful new ventures in the field. Time and time again the rapid growth of the market has found existing companies too busy expanding markets or product lines to which they were already committed to explore some of the more speculative new markets or technologies. And so the door was left open for new ventures, typically headed (originally at least) by an entrepreneur with a research or marketing background who had enough faith in the new market and technology to gamble. Fortunately for the US economy capital was readily available in the late 1950s and the 1960s to finance these ventures, and approximately a hundred new companies were formed to produce semiconductor devices; many of them made significant contributions to the development of microelectronics. Two such contributions in which I was directly involved were the development of the planar transistor and planar integrated circuit at Fairchild when that organisation was only two years old, and the development of the microprocessor at the Intel Corporation only two years after that company was founded. There are many more examples. The environment for entrepreneurial innovation in the US is not matched in other industrialized nations and it has been a major contributor to America's leadership in the field.

The growth of microelectronics has in turn created other opportunities. A host of companies have been established to serve the needs of the integrated-circuit producers. These companies supply everything from single-crystal silicon to computer-controlled design aids to automatic test equipment and special tooling. Often the novel consumer products that are spawned by developments in microelectronics have been manufactured and marketed initially by new companies. The digital watch and the television game are familiar examples.

When the integrated circuit was still an infant, Patrick E. Haggerty of Texas Instruments called attention to the increasing pervasiveness of electronics and predicted that electronic techniques would continue to displace other modes of control, reaching into nearly all aspects of our lives. Just such a displacement has been taking place, primarily because the microelectronics industry has been able to make ever more sophisticated functional elements at ever decreasing costs. Mechanical elements of the calculator and the watch have been displaced by integrated circuits that are less expensive and also offer more flexibility. Now the electromechanical functions of vending machines, pinball machines and traffic signals are being displaced. In the near future the automobile

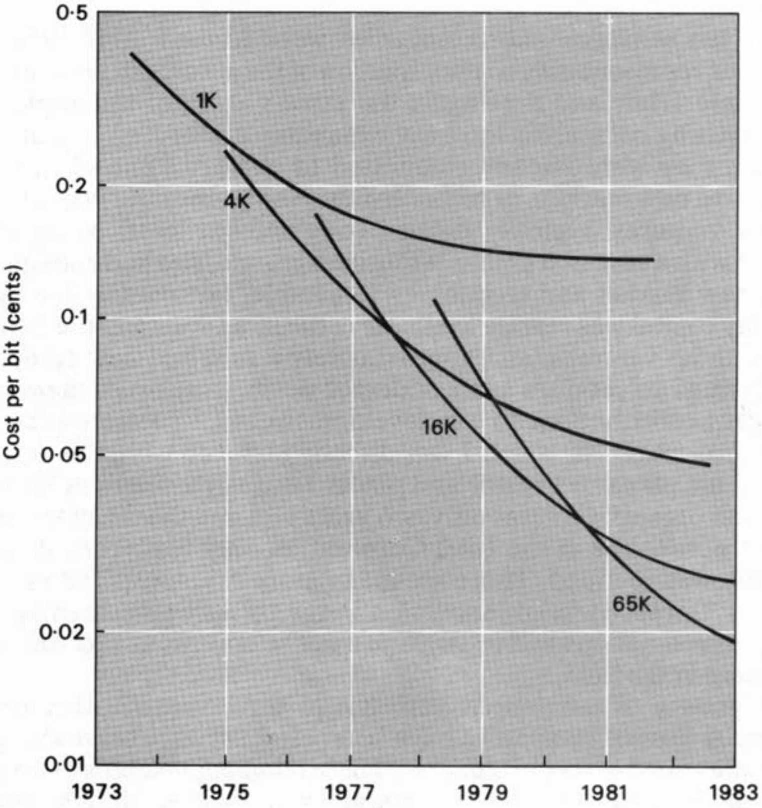


Fig. 1.3 Cost per bit of computer memory

engine will be controlled by a computer, with a consequent improvement in efficiency and reduction of pollutants. All these applications are simply extensions of the traditional applications of electronics to the task of handling information in measurement, communication and data manipulation. It has often been said that just as the industrial revolution enabled man to apply and control greater physical power than his own muscle could provide, so electronics has extended his intellectual power. Microelectronics extends that power still further.

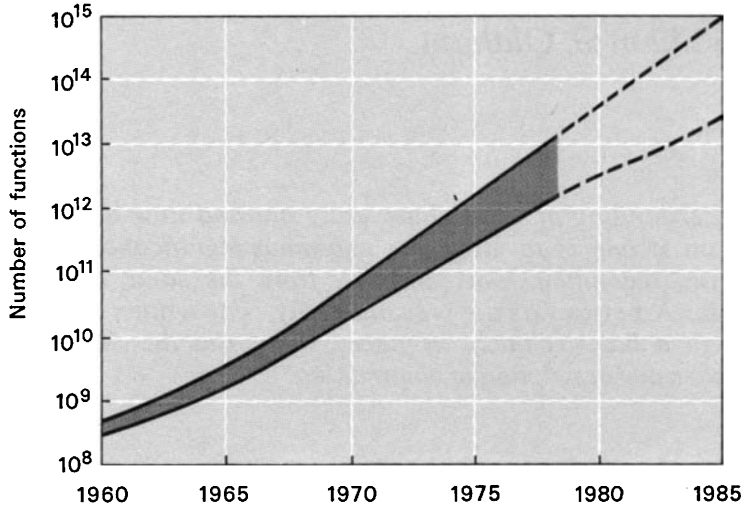


Fig. 1.4 Annual utilization of electronic functions

By 1986 the number of electronic functions incorporated into a wide range of products each year can be expected to be one hundred times greater than it is today. The experience curve predicts that the cost per function will have declined by then to a twentieth of the 1976 cost, a reduction of twenty-five per cent per year. At such prices electronic devices will be exploited even more widely, augmenting mail service, expanding the library and making its contents more accessible, providing entertainment, disseminating knowledge for educational purposes and performing many more of the routine tasks in the home and office. It is in the exponential proliferation of products and services dependent on microelectronics that the real microelectronic revolution will be manifested.

The Fabrication of Microelectronic Circuits

William G. Oldham

An understanding of what chips really are and how they are made is important if one is to grasp the enormous significance of the microelectronics revolution. This account, from the same special issue of Scientific American *as the previous essay, was written for the layman. Although a little technical in places, it remains the clearest and most comprehensive description of chipmaking.*

The manufacture of large-scale integrated circuits has as its primary goal the lowest possible cost per electronic function performed. The main features of the fabrication processes adopted by the microelectronics industry can be best understood in terms of this goal. These features include the fabrication of many circuits at a time (an extraordinary example of mass production), the reduction of the circuits to the smallest possible size and the maximum simplification of the processing technology.

The dramatic reduction in the cost of microelectronic circuits achieved in the past few years has not resulted from any major new breakthrough in fabrication technology. Indeed, most of the basic manufacturing processes involved have been widely adopted in the industry for five years or more. The recent sharp drop in fabrication cost has been achieved during a period of general economic inflation. The cost of processing a "wafer" of silicon, the substrate on which the microelectronic circuits are made, has risen moderately, but the area of the wafers has increased more rapidly, approximately doubling every four years. Thus the processing cost per unit area has actually decreased. Meanwhile the space required for a given electronic function has shrunk by a factor of two every eighteen months or so. This reduction in size has come not only from great ingenuity in designing simple circuits and simple technological processes for making them but also from the continuing miniaturization of the circuit elements and their interconnections. Moreover, the gradual elimination of defects in various manufacturing steps has resulted in a significant decrease in the net cost of fabrication. With a lower frequency of defects the yield of good circuits on a given wafer increases.

The current pace of developments in the manufacturing of microelectronic circuits suggests that the progress in reducing production costs will continue. Virtually every stage of fabrication – from photolithography to packaging – is either in the midst of a significant advance or on the verge of one.

A large-scale integrated circuit contains tens of thousands of elements, yet each element is so small that the complete circuit is typically less than a quarter of an inch on a side. The pure, single-crystal silicon wafers that bear the circuits are much larger: currently three or four inches in diameter. One of the key economies in the manufacture of microelectronic circuits is the simultaneous fabrication of hundreds of circuits side by side on a single wafer. An even greater scale of mass production is attained in several stages of manufacture by processing as many as one hundred wafers together in a batch. Hence the cost of labour and equipment is shared by thousands of circuits, making possible the extremely low per-circuit cost that is characteristic of microelectronics.

After a wafer has passed through the fabrication stage (see Figure 1.5) it is sectioned into individual dice, or chips, each of which is a complete microelectronic circuit. Not all the circuits will work. Defects in a wafer cannot be avoided, and a single defect can ruin an entire circuit. For example, a scratch only a few micrometres long can break an electrical connection. It is impossible both physically and economically to repair the defective circuits; they are simply discarded.

As one might expect, the larger the die, the greater the chance for a defect to appear and render the circuit inoperative. The yield (the number of good circuits per wafer) decreases with the size of the dice, both because there are fewer places on a wafer for larger dice and because the larger circuits are more likely to incorporate a defect.

At first it might appear most economical to build very simple, and therefore very small, circuits on the grounds that more of them would be likely to be good ones. It is true that small circuits are inexpensive; simple logic circuits are available for as little as ten cents each. The costs of testing, packaging and assembling the completed circuits into an electronic system, however, must also be taken into account. Once the circuits are separated by breaking the wafer into dice, each die must be handled individually. From that point on the cost of any process such as packaging or testing is not shared by hundreds or thousands of circuits. In fact, for medium-scale integrated circuits packaging and testing costs often dominate the other production costs.

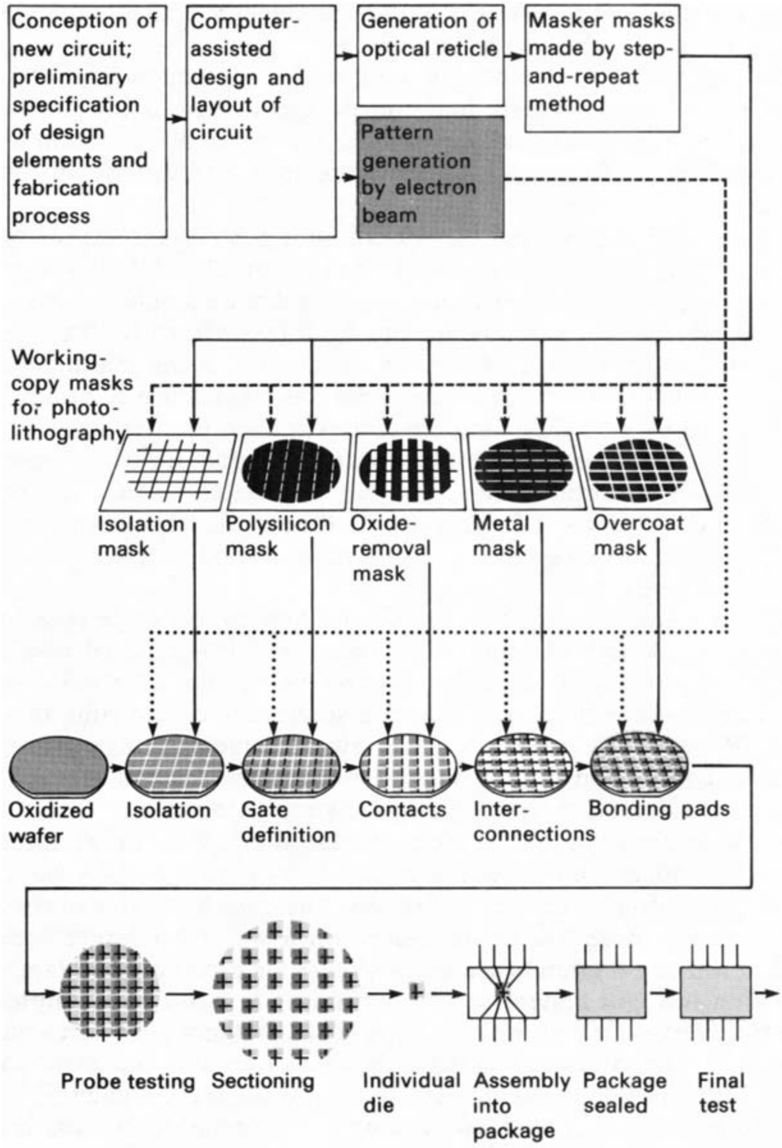


Fig. 1.5 An outline of the manufacture of a large-scale integrated circuit

A typical microelectronic system contains both large-scale and medium-scale integrated circuits, and the cost of designing and constructing the system rises rapidly as the number of circuits increases. To minimize the total cost of the system one would like ideally to use either a small number of very powerful circuits (requiring that each circuit be large) or a large number of very cheap circuits (requiring that each circuit be small). In many complex systems the minimum total cost is achieved with the more expensive, larger circuits: ones costing closer to $10 each than to $1. If less powerful, cheaper circuits were to be used, many more would be needed for constructing the system, and testing and assembly costs would tend to build up. On the other hand, if too much electronic function is packed into a circuit, the large size of the die would result in such a low yield per wafer that the cost per circuit would become prohibitive.

Assuming a typical selling price of $10 for a large-scale integrated circuit, one can work backward and estimate optimum die sizes. (The selling price of the circuit must of course be high enough to recover not only the direct costs of manufacture but also the costs of research and development, marketing and general overheads; thus it is reasonable to assume that the direct manufacturing cost is a good deal less, say about $5.) It costs roughly $100 just to process a silicon wafer, regardless of the size of the dice; when testing and packaging are included, the total manufacturing cost is perhaps doubled. Hence if the manufacturing cost of the particular integrated circuit in question is assumed to be $5, the optimum die size is one that yields about forty good circuits per wafer. At present it is possible to achieve such a yield for dice that measure approximately five millimetres on a side. It is interesting to note that in this example a rather low percentage of the circuits on the wafer are good. The yield is only forty good circuits out of the 250 that can be fabricated in a single wafer 100 millimetres in diameter.

The structure of an integrated circuit is complex both in the topography of its surface and in its internal composition. Each element of such a device has an intricate three-dimensional architecture that must be reproduced exactly in every circuit. The structure is made up of many layers, each of which is a detailed pattern. Some of the layers lie within the silicon wafer and others are stacked on the top. The manufacturing process consists in forming this sequence of layers precisely in accordance with the plan of the circuit designer.

Before examining how these layers are formed, it will be helpful to take an overall look at the procedure by which an integrated circuit is transformed from a conception of the circuit designer to a physical reality. In the first stage of the development of a new microelectronic circuit the designers who conceive of the new product work at specifying the functional characteristics of the device. They also select the processing steps that will be required to manufacture it. In the next stage the actual design of the device begins: the size and approximate location of every circuit element are estimated. Much of this preliminary design work is done with the aid of computers.

A computer can simulate the operation of the circuit in much the same way that electronic television games simulate the action of a table-tennis game or a space war. The circuit designer monitors the behaviour of the circuit voltages and adjusts the circuit elements until the desired behaviour is achieved. Computer simulation is less expensive than assembling and testing a "breadboard" circuit made up of discrete circuit elements; it is also more accurate. The main advantage of simulation, however, lies in the fact that the designer can change a circuit element merely by typing in a correction on a keyboard, and he can immediately observe the effect of the modification on the behaviour of the circuit.

The final layout giving the precise positions of the various circuit elements is also made with the aid of a computer. The layout designer works at a computer terminal, placing and moving the circuit elements while observing the layout magnified several hundred times on a cathode-ray-tube display. The layout specifies the pattern of each layer of the integrated circuit. The goal of the layout is to achieve the desired function of each circuit in the smallest possible space. The older method of drawing circuit layouts by hand has not been entirely replaced by the computer. Many parts of a large-scale integrated circuit are still drawn by hand before being submitted to the computer.

At each stage of this process, including the final stage when the entire circuit is completed, the layout is checked by means of detailed computer-drawn plots. Since the individual circuit elements can be as small as a few micrometres across, the checking points must be greatly magnified; usually the plots are 500 times larger than the final size of the circuit.

The time required to complete the task of circuit design and layout varies greatly with the nature of the circuit. The most difficult circuits to design are microprocessors, and here the design and layout can take several years. Other devices, such as static memories with a largely repetitive pattern, can be designed and laid out more quickly, in some cases in only a few months.

When the design and layout of a new circuit is complete, the computer memory contains a list of the exact position of every element in the circuit. From that description in the computer memory a set of plates, called photomasks, is prepared. Each mask holds the pattern for a single layer of the circuit. Since the circuits are so small, many can be fabricated side by side simultaneously on a single wafer of silicon. Thus each photomask, typically a glass plate about five inches on a side, has a single pattern repeated many times over its surface.

The manufacture of the photomasks is an interesting story in itself. Typically the process consists in first generating from the computer memory a complete pattern for each layer of the circuit. That is done by scanning a computer-controlled light spot across a photographic plate in the appropriate pattern. This primary pattern, called the reticle, is checked for errors and corrected or regenerated until it is perfect. Typically the reticle is ten times the final size of the circuit. An image of the reticle that is one tenth its original size is then projected optically on the final mask. The image is reproduced side by side hundreds of times in a process called "step and repeat". The constraints on both the mechanical system and the photographic system are demanding; each element must be correct in size and position to within about one micrometre. The original plate created by the step-and-repeat camera is copied by direct contact printing to produce a series of submasters. Each submaster serves in turn to produce a large number of replicas, called working plates, that will serve for the actual fabrication process. The working plate may be either a fixed image in an ordinary photographic emulsion or a much more durable pattern etched in a chromium film on a glass substrate.

A complete set of correct masks is the culmination of the design phase of the development of the microelectronic circuit. The plates are delivered to the wafer-fabrication facility, where they will be used to produce the desired sequence of patterns in a physical structure. This manufacturing facility receives silicon wafers, process chemicals and photomasks. A typical small facility employing a hundred people can process several thousand wafers per week. Assuming that there are fifty working circuits per wafer, such a plant can produce five million circuits per year.

The inside of a wafer-fabrication facility must be extremely clean and orderly. Because of the smallness of the structures being manufactured even the tiniest dust particles cannot be tolerated. A single dust particle can cause a defect that will result in the malfunction of a circuit. Special clothing is worn to protect the manufacturing environment from dust carried by the human operators. The air is continuously filtered and recirculated to keep the dust level at a minimum. Counting all the dust particles that are a micrometre or more in diameter, a typical wafer-fabrication plant harbours fewer than one hundred particles per cubic foot. For the purpose of comparison, the dust level in a modern hospital is on the order of 10,000 particles per cubic foot.

The circuit manufacturer often buys prepared wafers of silicon ready for the first manufacturing step. The low price (less than $10) of a prepared wafer belies the difficulties encountered in its manufacture. Raw silicon is first reduced from its oxide, the main constituent of common sand. A series of chemical steps are taken to purify it until the purity level reaches 99.9999999 per cent. A charge of purified silicon, say ten kilograms, is placed in a crucible and brought up to the the melting point of silicon: 1420 degrees Celsius. It is necessary to maintain an atmosphere of purified inert gas over the silicon while it is melted, both to prevent oxidation and to keep out unwanted impurities. The desired impurities, known as dopants, are added to the silicon to produce a specific type of conductivity, characterized by either positive (*p* type) charge carriers or negative (*n* type) ones.

A large single crystal is grown from the melt by inserting a perfect single-crystal "seed" and slowly turning and withdrawing it. Single crystals three to four inches in diameter and several feet long can be pulled from the melt. The uneven surface of a crystal as it is grown is ground to produce a cylinder of standard diameter, typically either three inches or one hundred millimetres (about four inches). The crystal is mounted in a fixture and cut into wafers with a thin high-speed diamond saw. In the finishing step the wafers are first smoothed on both sides by grinding and then are highly polished on one side. The final wafer is typically about half a millimetre thick. The final steps must also be carried out in an absolutely clean environment. There can be no defects, polishing damage, scratches or even chemical impurities on the finished surface.

The dominant role of silicon as the material for microelectronic circuits is attributable in large part to the properties of its oxide. Silicon dioxide is a clear glass with a softening point higher than 1400 degrees C. It plays a major role both in the fabrication of silicon devices and in their operation. If a wafer of silicon is heated in an atmosphere of oxygen or water vapour, a film of silicon dioxide forms on its surface. The film is hard and durable and adheres well. It makes an excellent insulator. The silicon dioxide is particularly important in the fabrication of integrated circuits because it can act as a mask for the selective introduction of dopants. Convenient thicknesses of silicon dioxide can be grown at temperatures in the range between 1000 and 1200 degrees C. The exact thickness can be accurately controlled by selecting the appropriate time and temperature of oxidation. For example, a layer of oxide one tenth of a micrometre thick will grow in one hour at a temperature of 1050 degrees C in an atmosphere of pure oxygen. A layer five times thicker will grow in the same time and at the same temperature in steam.

An important aspect of the oxidation process is its low cost. Several hundred wafers can be oxidized simultaneously in a single operation. The wafers are loaded into slots in a quartz "boat", separated by only a few millimetres. The high-temperature furnace has a cylindrical heating element surrounding a long quartz tube. A purified stream of an oxygen-containing gas passes through the tube. The boats of wafers are loaded into the open end and slowly pushed into the hottest part of the furnace. The temperature in the process zone is controlled to an accuracy of better than one degree C. Often the entire procedure is supervised by a computer. A small process-control computer monitors the temperature, directs the insertion and withdrawal of the wafers and controls the internal environment of the furnace.

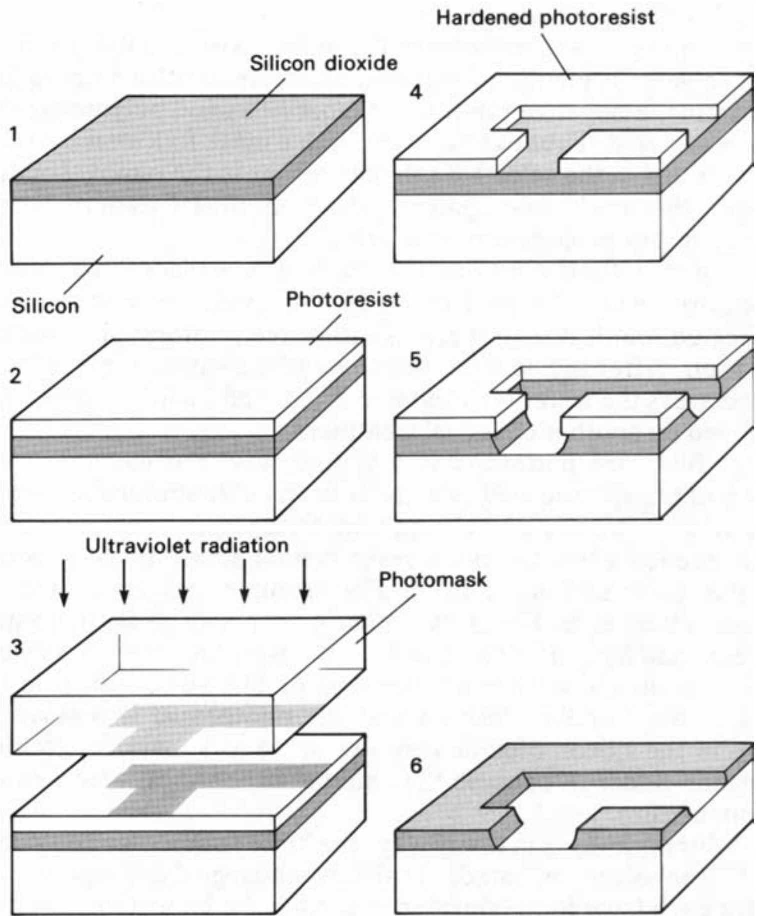


Fig. 1.6 The process of photolithography

The fabrication of integrated circuits requires a method for accurately forming patterns on the wafer. The microelectronic circuit is built up layer by layer, each layer receiving a pattern from a mask prescribed in the circuit design. The photoengraving process known as photolithography, or simply masking, is employed for the purpose.

The most basic masking step involves the etching of a pattern into an oxide. An oxidized wafer is first coated with photoresist, a light-sensitive polymeric material. The coating is laid down by placing a drop of the photoresist dissolved in a solvent on the wafer and then rapidly spinning the wafer. A thin liquid film spreads over the surface and the solvent evaporates, leaving the polymeric film. A mild heat treatment is given to dry out the film thoroughly and to enhance its adhesion to the silicon dioxide layer under it.

The most important property of the photoresist is that its solubility in certain solvents is greatly affected by exposure to ultraviolet radiation. For example, a negative photoresist cross-links and polymerizes wherever it is exposed. Thus exposure through a mask followed by development (washing in the selective solvent) results in the removal of the film wherever the mask was opaque. The photoresist pattern is further hardened after development by heating.

The wafer, with its photoresist pattern, is now placed in a solution of hydrofluoric acid. The acid dissolves the oxide layer wherever it is unprotected, but it does not attack either the photoresist or the silicon wafer itself. After the acid has removed all the silicon dioxide from the exposed areas the wafer is rinsed and dried, and the photoresist pattern is removed by another chemical treatment.

Other films are patterned in a similar way. For example, a warm solution of phosphoric acid selectively attacks aluminium and therefore can serve to pattern an aluminium film. Often an intermediate masking layer is needed when the photoresist cannot stand up to the attack of some particular etching solution. For example, polycrystalline silicon films are often etched in a particularly corrosive mixture containing nitric acid and hydrofluoric acid. In this case a film of silicon dioxide is first grown on the polycrystalline silicon. The silicon dioxide is patterned in the standard fashion and the photoresist is removed. The pattern in the silicon dioxide can now serve as a mask for etching the silicon film under it, because the oxide is attacked only very slowly by the acid mixture.

Photolithography is in many ways the key to microelectronic technology. It is involved repeatedly in the processing of any device, at least once for each layer in the finished structure. An important requirement of the lithographic process is that each pattern be positioned accurately with respect to the layers under them. One technique is to hold the mask

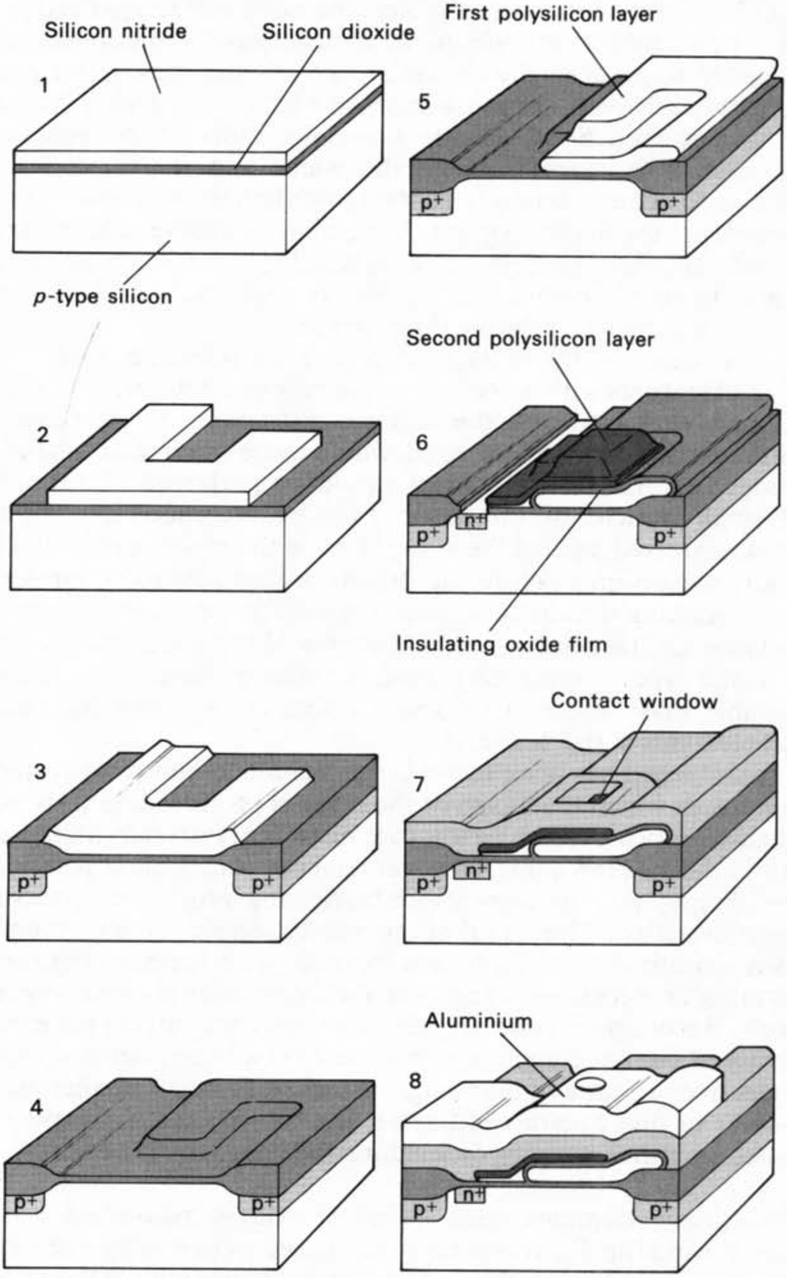


Fig. 1.7 The complete fabrication sequence for a two-level *n*-channel polysilicon-gate MOS circuit element

just off the surface and to visually align the mask with the patterns in the wafer. The machine that holds the wafer and mask for this operation can be adjusted to an accuracy of one or two micrometres. After perfect alignment is achieved, the mask is pressed into contact with the wafer. The mask is then flooded with ultraviolet radiation to expose the photoresist. The space between the wafer and the mask is often evacuated to achieve intimate contact; atmospheric pressure squeezes the wafer and the mask together. According to whether a high vacuum or a moderate one is used, the process is called "hard" or "soft" contact printing. In another variation, proximity printing, the mask is held slightly above the wafer during the exposure.

The variations in the masking process arise from the need to print very small features with no defects in the pattern. If the mask were to be positioned very far from the surface, diffraction of the ultraviolet radiation passing through the mask would cause the smaller features to blur together. Thus hard contact would be preferred. On the other hand, small particles on the wafer or mask are abraded into the mask when it is pressed against the wafer. Hence the masks can be used for only a few exposures before the defects accumulate to an intolerable level. A masking technique is chosen that is appropriate to the particular technology. Depending on the flatness of the wafer and the mask, and on the type of mask employed, a technique is chosen that gives reasonable mask life and sufficient resolution to print the smallest circuit elements in the device.

A recent trend has been towards the technique known as projection alignment, in which the image of the mask is projected onto the wafer through an optical system. In this case mask life is virtually unlimited. It is only in the past few years, however, that optics capable of meeting the photolithographic requirements for fabricating integrated circuits have become available. The fact that the wafers increase in size every few years is a continuing problem, and the task of designing optics capable of forming an accurate image over the larger area is becoming more difficult. Recent projection aligners, however, circumvent the extreme difficulty of constructing a lens capable of resolving micrometre-sized features over an area of many square inches. A much smaller area, of the order of one square centimetre, is exposed, and the exposure is repeated by either stepping or scanning the image over the wafer.

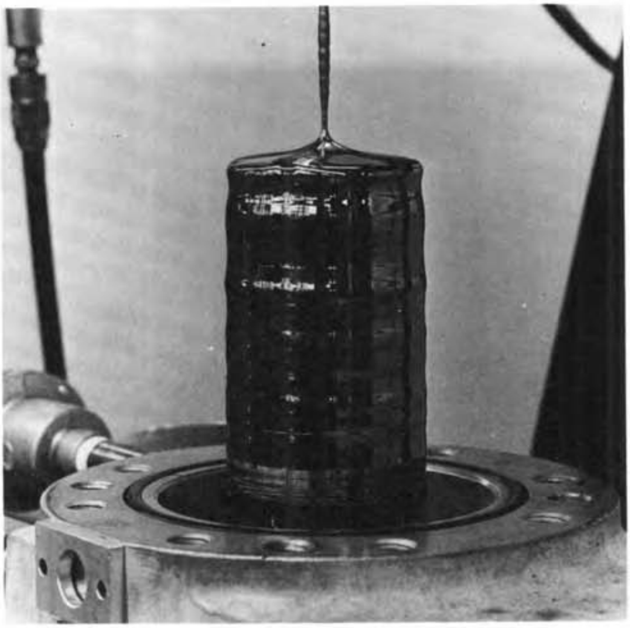
Active circuit elements such as metal-oxide-semiconductor (MOS) transistors and bipolar transistors are formed in part within the silicon substrate. To construct these elements it is necessary to selectively introduce impurities, that is, to create localized *n*-type and *p*-type regions by adding the appropriate dopant atoms. There are two

The microcircuit is laid out with the aid of a computer. Sections of the large-scale integrated circuit can be scrutinized and modified with a light pen on a visual display unit. *Photo courtesy of Applicon*

The chip layout is laboriously checked on a light table by the minute examination of each photomask image. The blow-up is 500 times normal size. *Photo by Jon Brenneis*

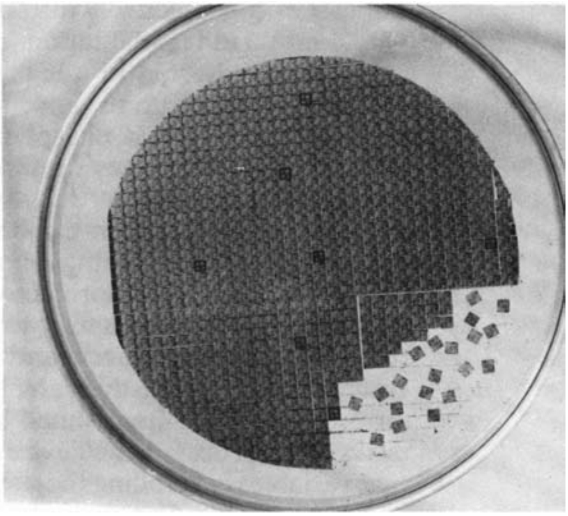
A typical wafer fabrication plant is kept spotlessly clean and the air continuously filtered to avoid even the tiniest of dust particles interfering with the manufacturing process. *Photo by Jon Brenneis*

This ingot of doped silicon is ready for slicing into thin wafers by a diamond-edged circular saw. *Photo by Jon Brenneis*

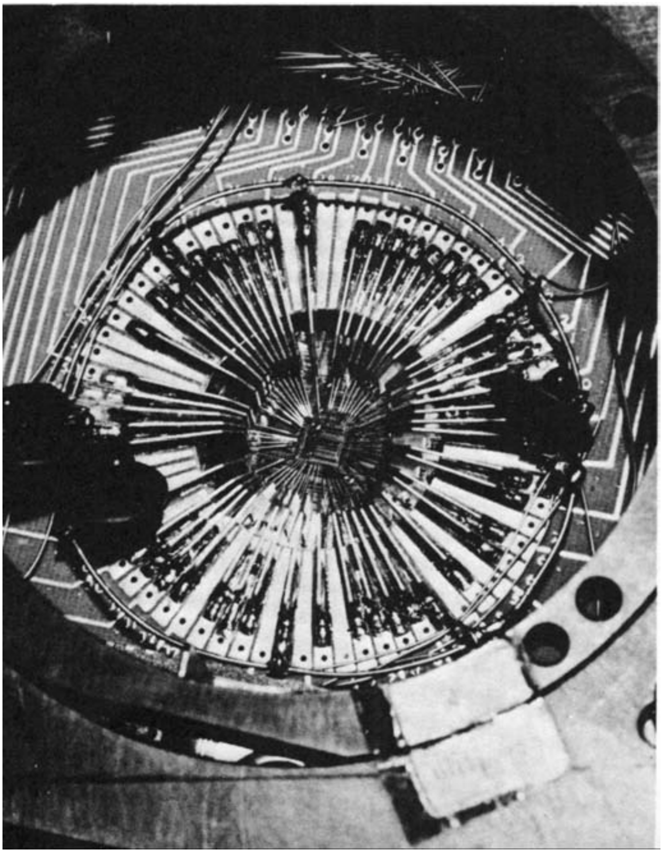


Once sliced, the wafers are sent in a glass "boat" for "cooking" in an oxidation furnace. *Photo courtesy of NCR Corporation*

When the chips have been built up on the wafer, grooves are cut into the surface, prior to breaking the wafer into individual dice. *Photo by Jon Brenneis*



Each of these dice contains a complete large-scale integrated circuit. Around two hundred are snapped off the average wafer. *Photo courtesy of British Information Services, Central Office of Information*



The chips are individually tested prior to packaging to see if their circuitry is defective. The long contact needles are directed on to the probe pads whilst viewing down a microscope. *Photo courtesy of Ferranti Electronics Limited* techniques for selectively introducing dopants into the silicon crystal: diffusion and ion implantation.

If silicon is heated to a high temperature, say 1000 degrees C, the impurity atoms begin to move slowly through the crystal. Certain key impurities (boron and phosphorus) move much more slowly through silicon dioxide than they do through silicon itself. This important fact enables one to employ thin oxide patterns as impurity masks. For example, a boat of wafers can be placed in a furnace at 1000 degrees in an atmosphere containing phosphorus. The phosphorus enters the silicon wherever it is unprotected, diffusing slowly into the bulk of the wafer. After enough impurity atoms have accumulated the wafers are removed from the furnace, and solid-state diffusion effectively ceases. Of course, every time the wafer is reheated the impurities again begin to diffuse; hence all the planned heat treatments must be considered in designing a process to achieve a specific depth of diffusion. The important variables controlling the depth to which impurities diffuse are time and temperature. For example, a layer of phosphorus one micrometre deep can be diffused in about an hour at 1100 degrees.

To achieve maximum control most diffusions are performed in two steps. The predeposit, or first, step takes place in a furnace whose temperature is selected to achieve the best control of the amount of impurity introduced. The temperature determines the solubility of the dopant in the silicon, just as the temperature of warm water determines the solubility of an impurity such as salt. After a comparatively short predeposit treatment the wafer is placed in a second furnace, usually at a higher temperature. This second heat treatment, the "diffusion drive-in" step, is selected to achieve the desired depth of diffusion.

In the formation of *pn* junctions by solid-state diffusion the impurities diffuse laterally under the oxide mask about the same distance as the depth of the junction. The edge of the *pn* junction is therefore protected by a layer of silicon dioxide. This is an important feature of the technique, because silicon dioxide is a nearly ideal insulator, and many of the electronic devices will not tolerate any leakage at the edge of the junction.

Another selective doping process, ion implantation, has been developed as a means of introducing impurities at room temperature. The dopant atoms are ionized (stripped of one or more of their electrons) and are accelerated to a high energy by passing them through a potential difference of tens of thousands of volts. At the end of their path they strike the silicon wafer and are embedded at various depths depending on their mass and their energy. The wafer can be selectively masked against the ions either by a patterned oxide layer, as in conventional diffusion, or by a photoresist pattern. For example, phosphorus ions

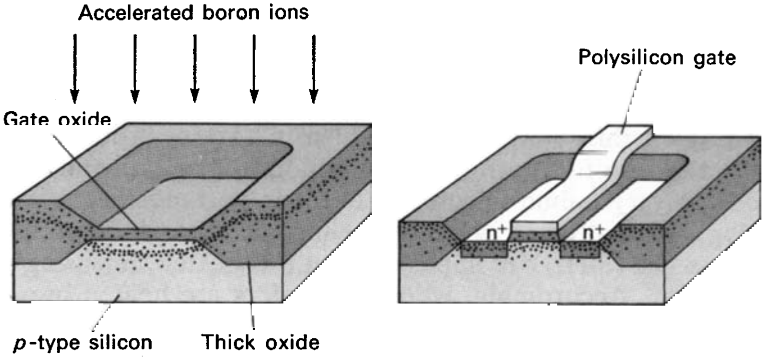


Fig. 1.8 Ion implantation

accelerated through a potential of 100,000 volts will penetrate the photoresist to a depth of less than half a micrometre. Wherever they strike bare silicon they penetrate to an average depth of a tenth of a micrometre. Thus even a one-micrometre layer of photoresist can serve as a mask for the selective implantation of phosphorus.

As the accelerated ions plough their way into the silicon crystal they cause considerable damage to the crystal lattice. It is possible to heal most of the damage, however, by annealing the crystal at a moderate temperature. Little diffusion takes place at the annealing temperature, so that the ion-implantation conditions can be chosen to obtain the desired distribution. For example, a very shallow, high concentration of dopant can be conveniently achieved by ion implantation. A more significant feature of the technique is the possibility of accurately controlling the concentration of the dopant. The ions bombarding the crystal each carry a charge, and by measuring the total charge that accumulates the number of impurities can be precisely determined. Hence ion implantation is used whenever the doping level must be very accurately controlled. Often ion implantation simply replaces the predeposit step of a diffusion process. Ion implantation is also used to introduce impurities that are difficult to predeposit from a high-temperature vapour. For example, the current exploration of the use of arsenic as a shallow *n*-type dopant in MOS devices coincides with the availability of suitable ion-implantation equipment.

A unique feature of ion implantation is its ability to introduce impurities through a thin oxide. This technique is particularly advantageous in adjusting the threshold voltage of MOS transistors. Either *n*-type or *p*-type dopants can be implanted through the gate oxide, resulting in either a decrease or an increase of the threshold voltage of the device. Thus by means of the ion implantation technique it is possible to fabricate several different types of MOS transistors on the same wafer.

The uppermost layers of integrated circuits are formed by depositing and patterning thin films. The two most important processes for the deposition of thin films are chemical-vapour deposition and evaporation. The polycrystalline silicon film in the important silicon-gate MOS technology is usually laid down by means of chemical-vapour deposition. Silane gas (SiH_4) decomposes when it is heated, releasing silicon and hydrogen. Accordingly, when the wafers are heated in a dilute atmosphere of silane, a uniform film of polycrystalline silicon slowly forms on the surface. In subsequent steps the film is doped, oxidized and patterned.

It is also possible to deposit insulating films such as silicon dioxide or silicon nitride by means of chemical-vapour deposition. If a source of oxygen such as carbon dioxide is present during the decomposition of silane, silicon dioxide is formed. Similarly, silicon nitride is grown by decomposing silane in the presence of a nitrogen compound such as ammonia.

Evaporation is perhaps the simplest method of all for depositing a thin film, and it is commonly employed to lay down the metallic conducting layer in most integrated circuits. The metallic charge to be evaporated, usually aluminium, is placed in a crucible, and the wafers to be coated are placed above the crucible in a movable fixture called a planetary. During evaporation the wafers are rotated in order to ensure the maximum uniformity of the layer. The motion of the planetary also wobbles the wafers with respect to the source in order to obtain a continuous aluminium film over the steps and bumps on the surface created by the preceding photolithographic steps. After a glass bell jar is lowered over the planetary device and a high vacuum is established the aluminium charge is heated by direct bombardment with high-energy electrons. A pure aluminium film, typically about a micrometre thick, is deposited on the wafer.

In the fabrication of a typical large-scale integrated circuit there are more thin-film steps than diffusion steps. Therefore thin-film technology is probably more critical to the overall yield and performance of the circuits than the diffusion and oxidation steps are. In a recent development a thin film is even employed to select the areas on a wafer that are to be oxidized. The compound silicon nitride has the property that it oxidizes much more slowly than silicon. A layer of silicon nitride can be vapour-deposited, patterned and used as an oxidation mask. The surface that results is much flatter than the surface if the thick oxide is grown everywhere and selectively removed. For *n*-channel MOS devices there is the additional advantage that an ion-implantation step involving boron can be added just before the oxidation step, relying on the nitride pattern as a mask. This procedure results in a heavily doped *p*-type region located precisely under the oxide, which acts as an obstacle to the formation of channels from adjacent elements in the device.

This "channel stopper" diffusion step is necessary in high-performance *n*-channel MOS technology. Without selective oxidation a special masking step would have to be added. The spacing between elements would then necessarily be larger; hence selective oxidation leads to greater circuit density. Bipolar integrated circuits also benefit greatly from the use of selective oxidation. By replacing the conventional diffused isolation with oxide isolation the space taken up by one bipolar transistor is reduced by more than a factor of four.

The wafer-fabrication phase of manufacture ends with an electrical test. Each die on the wafer is probed to determine whether it functions correctly. The defective dice are marked with an ink spot to indicate that they should be discarded. A computer-controlled testing machine quickly tests each circuit, steps to the next one and performs the inking without human intervention. It can also keep accurate statistics on the number of good circuits per wafer, their location and the relative incidence of various types of failure. Such information is helpful in finding new ways to improve the yield of good circuits.

The completed circuit must undergo one last operation: packaging. It must be placed in some kind of protective housing and have connections with the outside world. There are many types of packages, but all have in common the fact that they are much larger and stronger than the silicon dice themselves. First the wafer is sectioned to separate the individual chips, usually by simply scribing between the chips and breaking the wafer along the scribe lines. The good circuits are bonded into packages, and they are connected to the electrodes leading out of the package by fine wires. The package is then sealed, and the device is ready for final testing. The packaged circuit goes through an exhaustive series of electrical tests to make sure that it functions perfectly and will continue to do so reliably for many years.

After the individual chips are obtained from the wafer the cost per manufacturing step rises enormously. No longer is the cost shared among many circuits. Accordingly automatic handling during packaging and testing must be introduced wherever possible. The traditional cost-saving technique has been to employ less expensive overseas labour for the labour-intensive packaging operation. As the cost of overseas labour rises and improved packaging technology becomes available, overseas hand labour is gradually being supplanted by highly automated domestic assembly.

A number of advanced processing techniques are now under development. For example, the simple wet-etching process in which films of aluminium or polycrystalline silicon are selectively removed is already being supplanted by dry-etching processes. Polycrystalline silicon can be "plasma-etched" in electrically excited gas of carbon tetrafluoride molecules (Freon). A high-frequency electric discharge at low pressure breaks the Freon molecule down into a variety of ions and free radicals (such as atomic fluorine). The free radicals attack the film but do not react with the photoresist mask. In addition to being a more controllable method for the selective removal of silicon, plasma-etching promises to be much less harmful to the environment. Instead of yielding large quantities of corrosive acids, the reaction products are very small quantities of fluorine and fluorides of silicon, which are easily trapped from the output of the system.

The technology of photolithography, which was stable for about ten years, is also undergoing several changes. First, as projection lithography replaces contact lithography, the number of defects is decreasing. Second, the availability of masks of higher quality is steadily reducing the size and cost of microelectronic devices. Third, new methods of lithography are being developed that could result in a tenfold reduction in the size of individual circuit elements and a hundredfold reduction in circuit area.

The smallest features that can be formed by the conventional photolithographic process are ultimately limited by the wavelength of light. Present technology can routinely reproduce elements a few micrometres across, and it appears possible to reduce the smallest features to about one micrometre. Electronic beams and X-rays, however, have wavelengths measured in nanometres (thousandths of a micrometre) and smaller; hence they are capable of producing extremely fine features.

X-ray lithography is simply a form of contact photolithography in which soft X-rays are substituted for ultraviolet radiation. The X-ray technique does indeed offer high resolution; simple structures less than a tenth of a micrometre across have already been produced. Because the entire wafer is exposed the process is also potentially quick and cheap. There are still many unsolved problems, however. X-ray masks, which consist of a heavy metallic pattern on a thin membrane such as Mylar, are fragile and difficult to make. It is also hard to align the mask with respect to the pattern on the wafer. Because of the attenuation of X-rays in air, the wafer must be exposed in a vacuum or in an atmosphere of helium. Present-day X-ray sources are comparatively weak, and long exposures are required. As a result no commercial integrated circuits have yet been manufactured with the aid of X-ray lithography.

Electron-beam lithography is an older and maturer technology, having its basis in electron microscopy. Actually a system for electron-beam lithography is much like a scanning electron microscope. A fine beam of electrons scans the wafer to expose an electron-sensitive resist in the desired areas. Although impressive results have been demonstrated, the application of electron-beam lithography is limited by its present high cost. The machines are expensive (roughly $1m per machine), and because the electron beam must scan the wafer rather than exposing it all at once the time needed to put the pattern on the wafer is quite long. The rate of progress in this area is rapid, however, and more practical systems are clearly on the way.

Although electron-beam lithography is currently too expensive to be part of the wafer-fabrication process, it is already a routine production technique for the making of photolithographic masks. With the aid of the electron-beam method it is possible to eliminate two photographic-reduction steps and write the pattern directly on the mask from the information stored in the computer memory. Masks can thereby be created in a few hours after the design is finished. The advantages of higher resolution, simplicity of manufacture and shorter production time may well result in the complete conversion of the industry to electron-beam mask-making. Gradually, as the cost decreases, electron-beam lithography will be introduced directly into the fabrication of wafers, and a new generation of even more complex microelectronic circuits will be born.

Guide to Further Reading

General

Following on from *Fortune* magazine, *Newsweek* carried an introductory article on "Computers: A New Wave" on 23 February 1976. Kenneth Lamott wrote a very simple account of "The Impudent, Magical Silicon Chip" in the magazine *Horizon* in July 1977.

Apart from the *Science* and *Scientific American* collections, which are strongly recommended (the latter being better on the actual technology, the former better on the social implications), an excellent description of "The Computer Society" appeared in *Time* magazine on 20 February 1978.

In the US, *Popular Science* also had special issues on "Microprocessors" in March 1977, and on "Microelectronics" in January 1978. In Britain, *Design* contained an introductory article by James Woudhuysen, "What Shall We Do With Microprocessors?" in September 1977; *Management Today* carried an authoritative acount of "The Micro-Revolution" in March 1978 by Philip Hughes and Trevor Armstrong of the British software firm Logica; and the editor of this volume contributed "The Microelectronic Revolution" and "Society With Chips – And Without Jobs" in *New Society*, 9 and 16 November 1978. The *Financial Times* had a special supplement on "Microelectronics" on 29 March 1979, as did *The Economist* on 1 March 1980.

By far the most important single event – and the one that really sparked widespread interest in the new technology in Britain – was the screening by the BBC of Ed Goldwyn's TV documentary "Now The Chips Are Down" in April 1978. This film has been widely acclaimed and it became a BBC worldwide bestseller. A shortened version of the script appears in chapter 6 of this book.

More recent introductions to the microelectronics revolution are:

"The Microprocessor: A Revolution for Growth", *Business Week*, 19 March 1979, which is recommended.

Adam Osborne, *Running Wild: The Next Industrial Revolution* (Osborne, Berkeley, 1979). Authoritative.

Ray Curnow and Susan Curran, *The Silicon Factor: Living with the Microprocessor* (National Extension College, Cambridge, 1979).

W. H. Mayall, *The Challenge of the Chip* (HMSO for Science Museum, London, 1980). Profusely illustrated.

Technology

The simplest introduction to microcomputing can be found in Adam Osborne, *An Introduction to Microcomputers* (Osborne, Berkeley, 1977), which consists of three volumes, Vol. 0, *The Beginner's Book*, Vol. 1, *Basic Concepts*, and Vol. 2, *Some Real Products*.

The best textbooks are generally reckoned to be:

Edwin E. Klingman, *Microprocessor Systems Design* (Prentice-Hall, Englewood Cliffs, New Jersey, 1977).

John B. Peatman, *Microcomputer-Based Design* (McGraw-Hill, New York, 1977).

But the following are widely used:

D. D. Givone and R. P. Roesser, *Microprocessors/Microcomputers: An Introduction* (McGraw-Hill, New York, 1980).

Daniel R. McGlynn, *Microprocessors: Technology, Architecture and Applications* (John Wiley, New York, 1976).

Carol Anne Ogdin, *Microcomputer Design* (Prentice-Hall, Englewood Cliffs, New Jersey, 1978).

G. W. Rao, *Microprocessors and Microcomputer Systems* (Van Nostrand Reinhold, 1978).

D. Roddy, *Introduction to Microelectronics* (Ontario, Canada, 1978).

Dwight H. Sawin, *Microprocessors and Microcomputer Systems* (Lexington Books, Lexington, Mass., 1977).

Charles J. Sippl, *The Microcomputer Handbook* (Petrocelli/Charter, New York, 1977).

M. E. Sloan, *Introduction to Minicomputers and Microcomputers* (Addison Wesley, London, 1980).

J. Watson, *Semiconductor Circuit Design* (Adam Hilger, Bristol, 1977).

Edward S. Young, *Fundamentals of Semiconductor Devices* (McGraw-Hill, New York, 1978).

In addition, there are two books in the *Electronics* magazine book series, *Microprocessors* and *Large-Scale Integration* (both McGraw-Hill, New York, 1976). "Primers" on the microprocessor have also appeared in the personal computer magazines, *Creative Computing* (Sept./Oct. 1977) and *Personal Computing* (June 1978).

A good technical account of the microprocessor by the co-founder of Intel can be found in Gordon E. Moore, "Microprocessors and Integrated Electronic Technology", *Proceedings of the IEEE*, Vol. 64, No. 6 (June 1976). Another neat introduction is "The Microprocessor In Control", a paper delivered to the Institution of Electrical Engineers in London on 10 October 1978, by H. A. Barker.

Christine Sutton, "Taking The Mystery Out Of Micro", *New Scientist*, 17 May 1979, explains how junction transistors, semiconductors and bubble memories actually work.

Infotech, *Microelectronics* (Maidenhead, UK, and Auerbach Publications, Pennsauken, New Jersey, 1980). The 58th Infotech state of the art report. This 650-page two-volume study is notable for the prediction that gallium arsenide will soon displace silicon as the base component of microprocessors.