Chapter 2
Computer Engineering and Nanotechnology

> *In those days computers were distributed by government order. Korolev and Mishin personally, wherever they could, attempted to attain the delivery of computers to Special Design Bureau number 1*
> Boris E. Chertok "Rockets and People. Race to the moon"
> *The most powerful experimental supercomputers in 1998, composed of thousands or tens of thousands of the fastest microprocessors and costing tens of millions of dollars, can do a few million MIPS. They are within striking distance of being powerful enough to match human brainpower, but are unlikely to be applied to that end. Why tie up a rare twenty-million-dollar asset to develop one ersatz-human, when millions of inexpensive original-model humans are available? Such machines are needed for high-value scientific calculations, mostly physical simulations, having no cheaper substitutes. AI research must wait for the power to become more affordable.*
> Hans Moravec "When will computer hardware match the human brain?"

2.1 A Brief History of Computing

When people began to understand themselves as reasonable beings, they felt the need to describe the world around them—to count everything their eye caught. The choice of the number system was quite natural. Ten fingers on the human hands, ten toes on the feet—such was the decisive argument in favor of the intuitively chosen decimal system. It has remained that way up to now. The symbols denoting numbers changed from time to time, but the system itself, most psychologically acceptable for us, remained the same practically everywhere.

It seems that very soon fingers and toes ceased to be sufficient, and the question of tools to facilitate calculations came up. The progress of computer technology is a century-long process which involved ingenious representatives of the human society (see the excellent reviews by B. N. Malinowski and B. A. Gladkikh). We know

N.G. Rambidi, *Molecular Computing*, DOI 10.1007/978-3-211-99699-7_2

that the great Leonardo da Vinci drew up plans for an adding machine which was reproduced in metal in our days and proved to be quite operational. This machine operated with 13 digit-registering wheels.

2.1.1 A Little Detail: The Mechanical Calculators

The basic idea of such a device is shown in Fig. 2.1. Suppose that each position of an arbitrary decimal number corresponds to a shaft on which two gear wheels with different numbers of operative teeth are mounted. Suppose that each shaft is set to its initial position corresponding to the number “0.” Other numbers correspond to subsequent rotations of the shaft by 36°. With a gear ratio between the gear wheels at adjacent shafts equal to 1:10, the shaft corresponding to units must spin 10 times by 36° to cause the shaft representing tens to move one position. This design allows dialing any number limited only by the number of shafts and to add to or subtract from it any number. Mechanical systems based on gear transmission were being perfected and used almost until the middle of the last century. Back in the 1950s the Soviet Army had in operational service the PUAZO antiaircraft artillery director which determined correction for velocity, humidity, temperature, etc., while controlling antiaircraft fire. PUAZO was a transportable cube with edges roughly one meter in length stuffed with drive gears, worm gears, electric motors, etc.

Over the centuries that followed the Leonardo da Vinci era, outstanding scholars from different countries—Blaise Pascal, Gottfried Leibniz, and Charles Babbage—occupied themselves with the development of computer technology. However, by the time the complex, cumbersome computations turned out to be vital, the level of computing technology was inadequate for addressing the most pressing challenges.

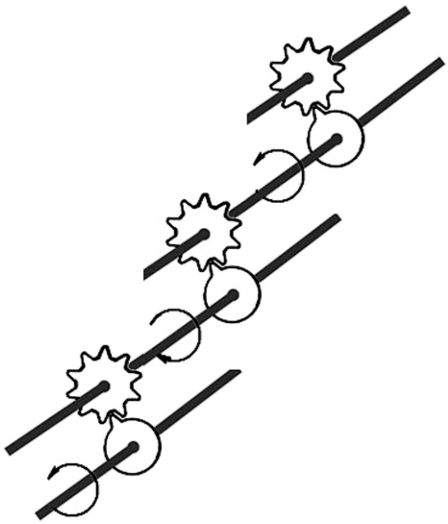


Fig. 2.1 Scheme of a mechanical computing device

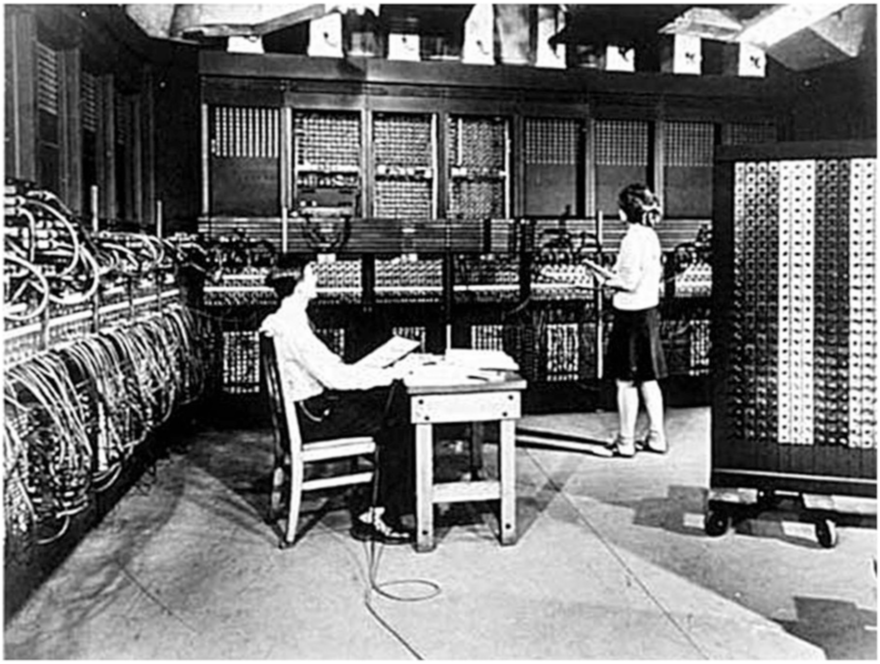


Fig. 2.2 Electronic computing device ENIAC—input of initial data

During the Second World War, improvement and development of new types of weapons generated overwhelming requirements to computing.

In 1941, the personnel of the Aberdeen Ballistic Research Laboratory in the United States approached the University of Pennsylvania's School of Electrical Engineering located nearby for help in creating firing tables. They suggested using the available Bush differential analyzer—a bulky mechanical analogous computing device. However, John Mauchly, a physicist at the School, proposed to create for this purpose a then powerful computer based on electronic valves. In April 1943 a contract was signed between the Ballistic Research Laboratory and the University of Pennsylvania to develop a computing machine, called the Electronic Numerical Integrator and Computer (ENIAC) with a budget of $400,000. About 200 people took part in this work, including several dozen mathematicians and engineers. The project was placed under the supervision of J. Mauchly and a talented electrical engineer Presper Eckert. The hard work was completed in late 1945 when ENIAC was successfully tested. In early 1946, the machine was first applied to solving real problems. Its dimensions were impressive: it measured 26 m in length, 6 m in height, and weighed 35 tons. It used the decimal numeral system and could hold in memory 20 ten-digit decimal numbers. ENIAC was programmed by being rewired via plugs and a patch panel which caused inconvenience since it could take many hours and even days to reprogram (Fig. 2.2). Therefore in 1945, while still completing the ENIAC project, its creators were already developing a new electronic

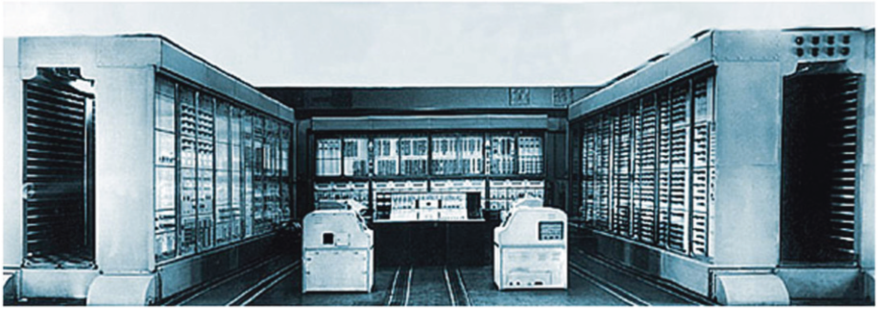


Fig. 2.3 First Soviet series-produced mainframe "Strela"

digital computer EDVAC. They suggested to store programs directly in the computer's random access memory, thus eliminating the main drawback of ENIAC—configuration using physical switches and plugs. At this time the world-renowned mathematician John von Neumann, member of the Manhattan project to develop the atomic bomb, joined the team. He immediately appreciated the perspectives of the new technology and took active part in creating EDVAC. The part of the report on the machine he wrote contained a general description of EDVAC and the fundamental principles of its design. In spite of the fact that a number of engineers were involved in the development of these principles, they later were called "von Neumann principles" (von Neumann paradigm). In any case, the establishment of these principles was a revolutionary event in computer technology that determined its further development.

Without going into detail of the history of the development of digital von Neumann computers, the most significant milestones are as follows.

In 1953 IBM began the production of a general-purpose computer. The series-produced IBM-701 had a random access memory with the capacity of 2,000 words of 36 bits each and was capable of ~10,000 operations per second.

Just 6 years later IBM-7030 installed at the Los Alamos Scientific Laboratory reached one million operations per second. Its random access memory capacity was 256,000 64-bit words.

The pace of progress in computer technology over the second half of the last century is illustrated by the characteristics of one of the most powerful supercomputers of our time, Top-500 Earth Simulator. It consists of 640 modules, each containing 8 processors. The system has 10 terabytes of RAM and is theoretically capable of 40 teraflops (Flop stands for the number of floating-point operations per second).

In the Soviet Union the first electronic computer was created in 1952 under the supervision of academician S. A. Lebedev. His first full-fledged computer, BESM, could perform 8,000 operations per second and store 1,000 39-bit words.

The first series-produced mainframe "Strela" was built in 1953 under the supervision of Yu. Ya. Bazilevsky (Fig. 2.3). Its speed was 2,000 operations per second, and its random access memory was based on cathode-ray storage tubes

(43-bit words). The machine occupied the area of ~300 m^2 without air conditioning which required almost as much space.

Very popular among computer specialists was Lebedev's BESM-6 which went into production in 1968. Its random access memory contained between 32,000 and 128,000 48-bit words, and its speed reached one million operations per second. Until 1987 355 such machines were manufactured.

A significant event in the history of computing was the introduction of personal computers, an entirely novel direction distinctly different from the general line of development in the 1940s–1970s. This has been facilitated by at least two factors. First, an urgent need was felt in different fields of human activity for computers that would be sufficiently powerful and at the same time simple in operation. Second, the electronic industry was essentially prepared for this need. A number of manufacturing companies brought to market electronic components, including microprocessors which were comparatively powerful for that time.

The first attempt to combine components into a single unit was made in 1975 by a company called MITS which released for sale a kit dubbed "Altair," essentially a set of parts and a housing. Buyers were supposed to solder together and test the assembled units and to create computer programs in machine language.

Subsequently a number of other versions of simple electronic computing devices appeared on the market. But the real beginning of the history of personal computers is associated with the names of American electronic engineers Steve Jobs and Steve Wozniak.

In the early 1970s Jobs worked at Atari where he met the senior developer of the company, Ron Wayne. Along with Wozniak, who at that time worked at Hewlett-Packard, they spent nights in a garage working in a personal computer which they called Apple-1. Soon after, on April 1, 1976, they created a company called Apple Computer. While establishing the company ran into significant difficulties. Constant need for funding, a cautious attitude of consumers to this new direction, and a number of other factors led to Ron Wayne's departure. Nevertheless, after the first model Apple-1 which received lukewarm response, Jobs and Wozniak created the Apple-2 computer which was a tremendous success. As a consequence, large companies started to produce personal computers. In the early 1980s, IBM has released its 5150 model, conspicuously called IBM PC, thus introducing the term "personal computer" widely used today.

It is now hard to find an area of human activity where these amazing computing devices would not be used for a broad range of tasks—from powerful computing to relaxing entertainment.

During the second half of the last century the information capacity of computer technology was steadily growing. A characteristic feature of this process was the preservation of the basic principles of computing devices and the continuous refinement of the components implementing these principles. In turn, progress in the component base was accompanied both by a change in the physical principles underlying the mechanisms of function of the components and by their consistent miniaturization, resulting in the need to change the production technology.

2.1.2 *Some Details: von Neumann Paradigm and Its Implementation*

The von Neumann paradigm encompasses six main principles:

1. Computers built on electronic elements should work in the binary rather than in the decimal system.
2. The program must be located in a storage device with sufficient capacity and appropriate read/write speed for commands of the program.
3. The program and the numbers with which the machine operates are represented in binary code. Thus, in terms of representation, commands and numbers are of the same type. This leads to the following important consequences:
   - Intermediate results of calculations, constants, and other numbers may be stored in the same storage device as the program.
   - Numerical notation of the program code allows the machine to perform operations on values that serve to encode program instructions.
4. The challenge of physical implementation of a mass storage device with a speed of operation corresponding to the performance of logic circuits requires a hierarchical organization of the memory.
5. Arithmetic units of the machine are constructed on the basis of circuits that perform the operation of addition. Implementing specialized devices for other operations is impractical.
6. The machine uses the principle of parallel operation on words, executed simultaneously over all bits. At the same time execution of program instructions (computer operations) is performed sequentially, one after another.

Today's computer technology is very diverse (Fig. 2.4). It includes digital computers in which program instructions, input data for solving problems, and computational results are recorded in memory as sets of binary characters. It also includes analog devices, processing continuous sequences of values of some physical quantity. Furthermore, specialized devices for mass solutions of a single task or a group of similar tasks also exist. However, the overwhelming majority of com-

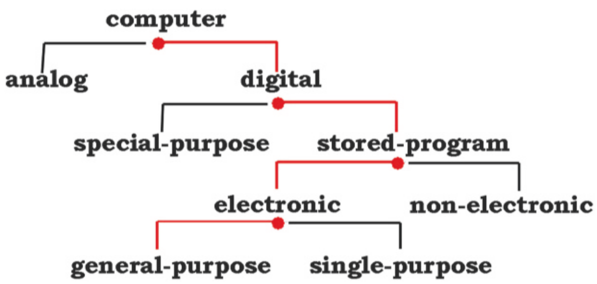


Fig. 2.4 Main directions of computer technology

putational devices that surround us in our daily life are universal digital computers. Let us consider the implementation of von Neumann principles on the example of a typical representative of this category of devices—a personal computer.

Computers of this type are built on the modular principle and represent a set of separate units. They communicate with each other using a special information highway—the bus. To create a flat bus multicore cables are usually used. The set of bus wires is divided into separate groups to transfer the address code of the operation to be performed, data, and control signals.

The principal computer components include (Fig. 2.5) a storage device, an arithmetic logic unit, and a control device.

The storage device, or memory, is a collection of cells designed to store information. Each cell is assigned a unique number called the address. Information stored in the cell may be both machine instructions and data. A machine instruction

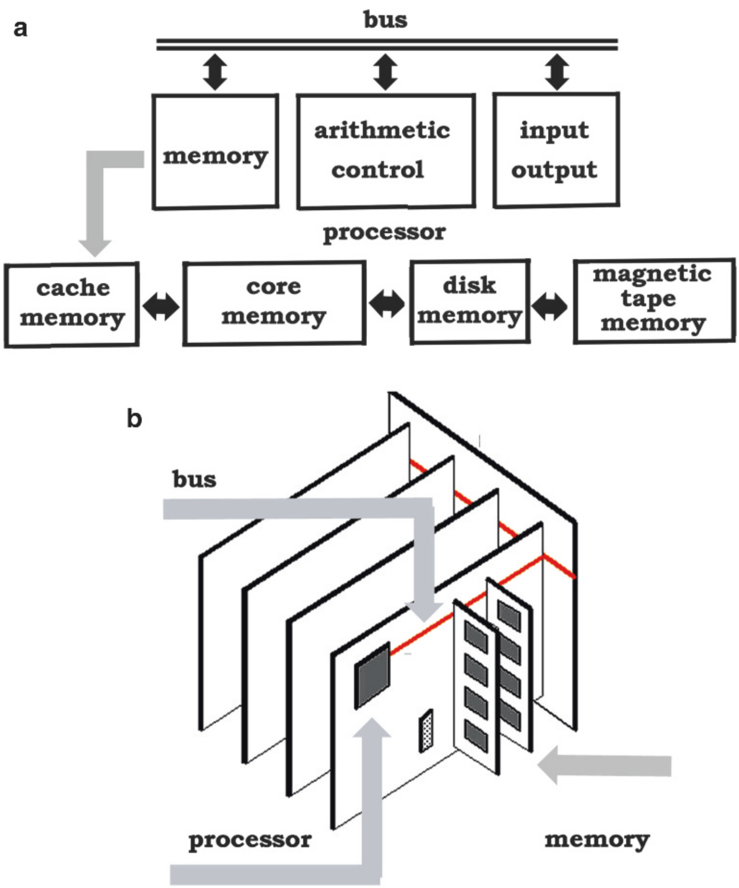


Fig. 2.5 Structure of a modern personal computer: (**a**) block scheme, (**b**) spatial location of devices

is a binary code that defines the operation to be carried out, the addresses of its operands (i.e., the codes of the numbers on which the operation is to be performed), and the address of the cell in which the result of the operation will be recorded. Storage devices of modern computers have a hierarchical structure (Fig. 2.5). The main memory is a solid state random access storage device with the read/write speed comparable with the processing speed of the arithmetic logic unit. In order to neutralize the difference in these speeds, an additional, small capacity high-speed memory can be used known as cache memory. In Pentium processors it contains 8,000 cells for code and another 8,000 cells for data. For technical and economic reasons the storage capacity of memory devices is limited. Today it reaches several gigabytes. Storage capacity can be extended by using slower storage on magnetic disks (up to hundreds of terabytes) and magnetic tapes with practically unlimited capacity.

All operations in a computer are controlled by the signals generated by the control unit. The control unit generates the address of the next command to be executed and sends a control signal for the contents of an appropriate memory cell to be read. The command readout is transmitted to the control unit. According to the information contained in the address fields of the command, the control unit generates the addresses of operands and control signals for reading the operands from the storage and transmitting them into the arithmetic logic unit. Subsequently the control unit sends signals for executing the operation to the arithmetic logic unit. The result is stored in the machine memory. Result attributes (sign, overflow flag, zero flag, etc.) are delivered to the controller where they are written in a status register. This information can be used while carrying out subsequent commands, e.g., conditional jump instructions.

2.2 Semiconductor Devices: A Revolution in Electronics

Tremendous progress made by computer technology is due to the development of the component base of computing devices. Over the past half century it underwent revolutionary changes that led to modern means of information processing utilized in virtually all areas of human life and activity.

The first major shift in digital computers was the transition from mechanical and relay systems to vacuum tubes. It is now even difficult to remember what vacuum tubes looked like and how radios, amplifiers, and control devices on their basis functioned. In the simplest case three electrodes—cathode, anode, and an intermediate electrode called the grid—were sealed into a glass vacuum tube. The cathode was heated by electric current and emitted electrons that were accelerated toward the anode by voltage passed through the grid. An electron tube, even in this simplest three-electrode implementation, is a natural embodiment of a switching element required for logical circuits. Depending on the potential on the grid, it either lets the current flow through or not, thus constituting an element with two stable states. Therefore, starting with ENIAC tube-based computing systems were created in different countries. However, thousands of electron tubes in a single device consumed a lot of energy and required labor-intensive maintenance by skilled personnel. Electron tube-based computing systems were capricious in operation. In order to achieve stable operation, days had to be spent on debugging the system. The temperature in the computer room was sometimes 10–15 °C. Therefore, the appearance of semiconductors dramatically increased the reliability of computers and paved the way for their further improvement.

It is not by accident that semiconductors won a strong position in computer engineering. As is known, in contrast to the quantum objects like atoms and molecules, solids have electronic band structure.

In order to understand the underlying reasons for the formation of the band structure, let us consider a simple model—a one-dimensional chain of atoms (Fig. 2.6). If there were two atoms, the electronic levels of such a system would be split into two components—a bonding and an antibonding orbital (see Chap. 3).

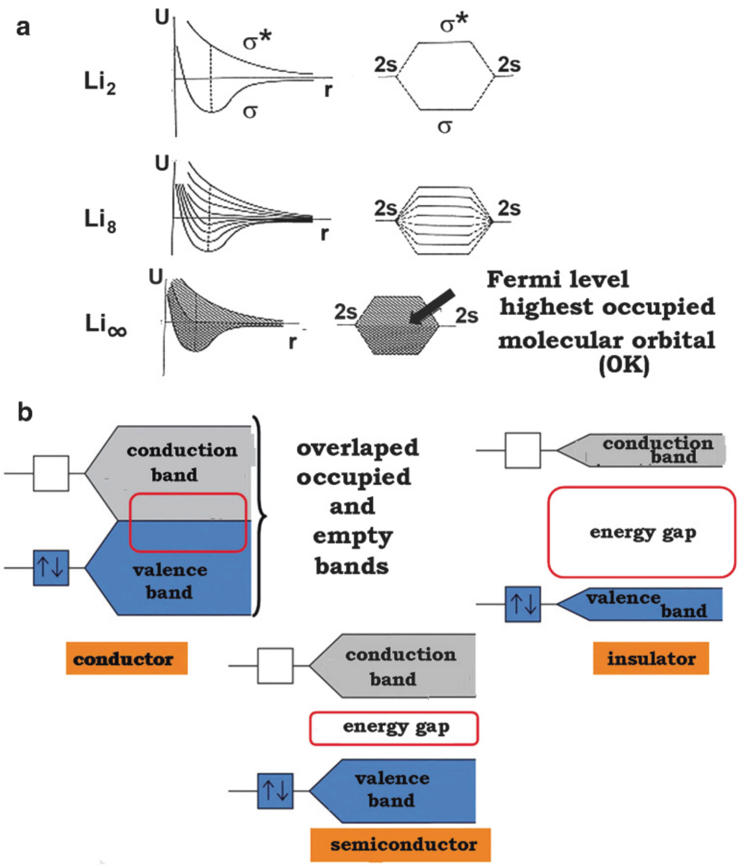


Fig. 2.6 Scheme of the formation of bands in a chain of Li atoms (**a**) and band structure of a conductor insulator and semiconductor (**b**)

As the number of atoms in the chain increases, so does the degree of level splitting and ultimately two energy bands arise which are called the valence band and the conduction band. They are continuous but contain a finite number of electronic states. For a number of elements these zones overlap (Fig. 2.6). The electrons of the atoms occupy the lower levels of the valence band. The other levels remain vacant and the electrons can move from the valence to the conduction band. Electronic conductivity arises in the system. Such electronic structure corresponds to conductors of electric current. In general the properties of solids are determined by the distance between the valence and conduction bands, i.e., by the band gap between them, and by the degree to which the valence band is filled by electrons. If the band gap is wide and the valence band is completely filled, the solid is an insulator. In the most interesting case, corresponding to a semiconductor, the valence band is nearly completely filled and the band gap is relatively narrow.

Semiconductors that practically do not contain dopants are called intrinsic, or undoped. In this case, when an excited electron is promoted from the valence to the conduction band, a positively charged vacancy is produced in the valence band. Of course the neighboring electrons can neutralize this vacancy, but while doing so they will form a new vacancy elsewhere. Thus, a positively charged moving entity appears in the semiconductor that is called a hole. In intrinsic semiconductors charge carriers must appear in pairs (electron–hole pair). The situation changes significantly if a certain amount of dopants—alloying additions—is introduced into the semiconductor. We will consider silicon whose electronic structure and properties correspond to a semiconductor. Tetravalent silicon forms four covalent bonds with neighboring atoms. If trivalent boron is introduced into the structure of the silicon crystal, one of the bonds remains unfilled (Fig. 2.7). It can be filled by an electron of any other neighboring silicon atom, leading to the formation of a hole. Dopants of this kind are called acceptors, and the resulting holes are situated just above the valence band. Such semiconductors are called p-type semiconductors. A different situation arises when a pentavalent atom of a dopant (phosphorus or antimony) is introduced into a silicon semiconductor. These atoms have five valence electrons, one more than silicon. The fifth electron is easily detached from the atom containing it. As a result a static ionic charge as well as an energy state corresponding to the fifth electron, situated slightly below the conduction band, arise. Such dopants are called donor dopants and the semiconductors n-type semiconductors.

Remarkable properties arise upon contact of p- and n-type semiconductors. In this case, due to large difference of concentrations—a whole sea of holes on one side and a sea of electrons on another side—strong diffusion currents of holes and electrons arise. As a result minority carriers appear in the system—electrons in p-type semiconductors and holes in n-type semiconductors. At the same time ions of the electron–hole pairs will be approaching the junction causing electric field around it (Fig. 2.8). In turn, this field will cause drift currents of electrons toward the n-type material and of holes toward the p-type material. If the external voltage is absent, the diffusion and drift currents will be equal in magnitude and opposite in direction. Therefore, the total current will be zero. Suppose that electric field is

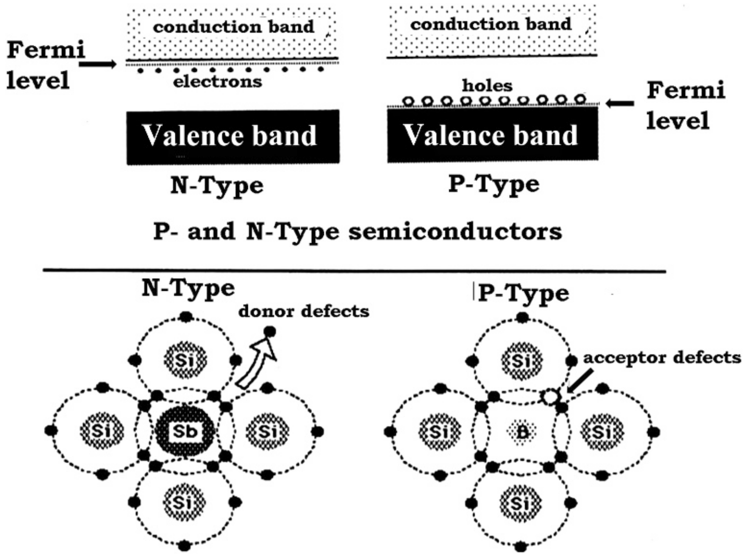


Fig. 2.7 Semiconductors of n- and p-type

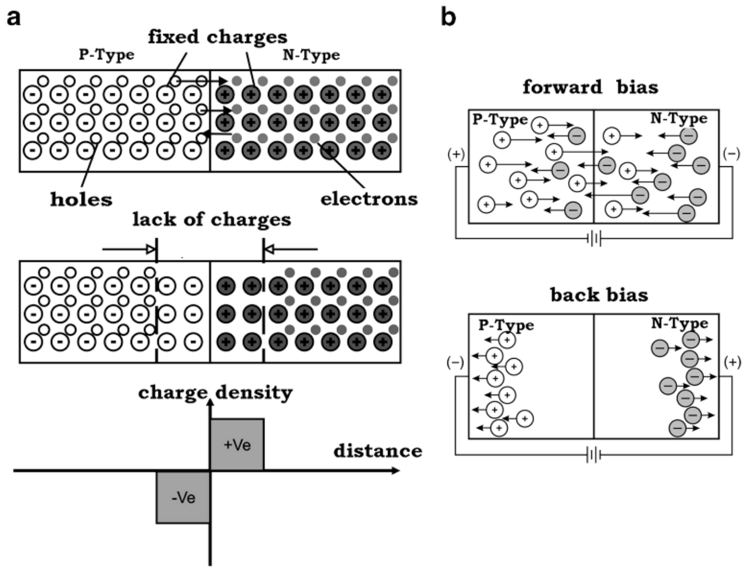


Fig. 2.8 p–n transition (**a**) and semiconductor diode (**b**)

applied to the junction (direct shift—a positive potential is applied to the n-type semiconductor and a negative one to the p-type semiconductor). This voltage will increase the concentration of minority carriers which will penetrate deeper into the material and recombine with majority carriers. Those minority carriers that

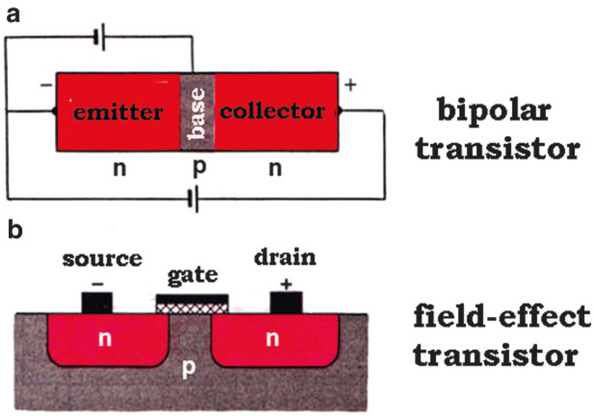
Fig. 2.9 First Bardeen and Brattain transistor

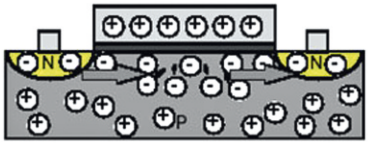
disappeared as a result of recombination will be replaced by new ones due to diffusion through the junction, leading to constant direct current. Similarly, if the field is applied in the reverse direction (reverse bias), the current through the junction will be vanishingly small (Fig. 2.9). Thus, the p–n transition behaves as a semiconductor diode.

All these properties of semiconductor materials and p–n and n–p junctions were used to create a unique semiconductor device—the transistor—which changed the face of computing devices. Today this name applies to a large group of semiconductor switching devices with two stable states.

In 1946 William Shockley, Walter Brattain, and John Bardeen at Bell Laboratories in New Jersey (AT&T Bell Labs) started to work on a semiconductor device—the transistor. In 1947 Bardeen and Brattain demonstrated the first implementation of a transistor on the basis of a germanium crystal with p- and n-zones, with metallic wires connected to the junction (Fig. 2.9, reference 6 in Chapter 2). Based on their work Shockley analyzed the physics of the device and a few months later proposed a fully planar semiconductor transistor. In 1956 Shockley, Bardeen, and Brattain received a Nobel Prize for this work. In his Nobel lecture John Bardeen said: "I knew the transistor was important, but I never foresaw the revolution in electronics it would bring."

Fig. 2.10 Bipolar (**a**) and field-effect (**b**) transistor





As a result of continued rapid development of the theory and semiconductor technology, various versions of two basic types of transistors—bipolar and field-effect transistors—were created (Fig. 2.10). They employ a combination of p–n and n–p semiconductor junctions.

A bipolar transistor comprises two semiconductor areas of the same type (emitter and collector), separated by a thin layer of semiconductor of the other type (base). In simple terms it can be thought of as two p–n junctions joined back-to-back. If voltage is applied only between the emitter and the collector, then for any polarity of the voltage current will not flow. One of the two p–n junctions will be closed. But when the voltage is applied between the emitter and the base, current flows in the chain emitter–collector, and the strength of this current can be controlled by much weaker emitter–base current.

The field-effect transistor introduced somewhat later is based on the idea expressed as early as 1925 by the American researcher Julius Lilienfeld. He proposed to control the resistance of a semiconductor layer in a system which is essentially a capacitor, with one plate made of metal and the second one from doped semiconductor, using voltage applied between the metal and the semiconductor. If negative potential is applied to the metallic plate, the field will displace the electrons from the surface layer of the semiconductor, leading to a lack of current carriers and an increase of resistance. When the polarity is reversed the number of carriers in this area will go up and the resistance will increase. The mechanism of action of the field-effect transistor is shown schematically in Figure.

A very important factor in creating the field-effect transistor was the availability of suitable materials: silicon (semiconductor) and silicon dioxide (insulator).

The latter can be easily grown on the silicon surface by oxidation. In this fashion the foundations of the modern MIS (metal–insulator–conductor) technology, also called MOS (metal–oxide–semiconductor) technology, were laid.

A revolutionary step in the development of the semiconductor technology was the transition to integrated circuits in which all the elements of the transistor are formed on the surface of a silicon crystal or some other semiconductor media. Their appearance was due to an acute need to improve the reliability of equipment and to automate manufacturing and assembly of electronic circuits. Assembling equipment at that time was mostly manual—a very laborious and time-consuming process poorly amenable to automation. As the number of switching devices in the digital equipment, especially in computers, increased manifold, the reliability and the mean time between failures dropped sharply. For example, the CD1604 computer released in 1960 by the US firm Control Data Corp. contained about 100,000 diodes and 25,000 transistors and could work without failure for not more than 2–3 h.

The world's first integrated circuits were designed and built in 1959 independently by Jack Kilby at Texas Instruments and Robert Noyce at the Fairchild Semiconductor Company.

In 1958 Kilby began to work on integrated circuits in which electronic components were supposed to be located on the same substrate. By this time semiconductor materials could be used to produce resistors, capacitors, and transistors. Resistors were produced using the ohmic properties of the semiconductor "body," while capacitors were built based on reverse biased p–n junctions.

In 1959 Kilby demonstrated the design of a flip-flop on one monolithic piece of germanium. For its production, photoengraving techniques patented by Texas Instruments were used. This "solid circuit" was introduced in 1960 at the exhibition organized by the American Institute of Radio Engineers.

Many of the shortcomings of "solid circuits" were later addressed by Robert Noyce, working at Fairchild. He developed technological processes that anticipated the modern semiconductor planar technology. A patent application was filed, and developers of components started to work on bringing diffusion resistors and transistors together on silicon wafers.

In 1960, a group of researchers at Fairchild Semiconductor headed by Jay Last produced the first integrated circuit containing four transistors (Fig. 2.11).

"You and I agree that while the world loves a hero, semiconductor progress depended on the efforts and ideas of a large number of people and that moving forward depended on contributions going back a few decades in some cases. Also, as is the case with most inventions, a number of people with access to the same pool of common knowledge were working independently at the same time to put it altogether and to make the necessary extensions to the existing technology and who realized that the time was right for society to accept the new concepts."—Jay Last later wrote in a letter to one of his friends.

The development of integrated circuits began to progress at a feverish pace. This was the beginning of a new era. To obtain a rough estimate of the rate of development,

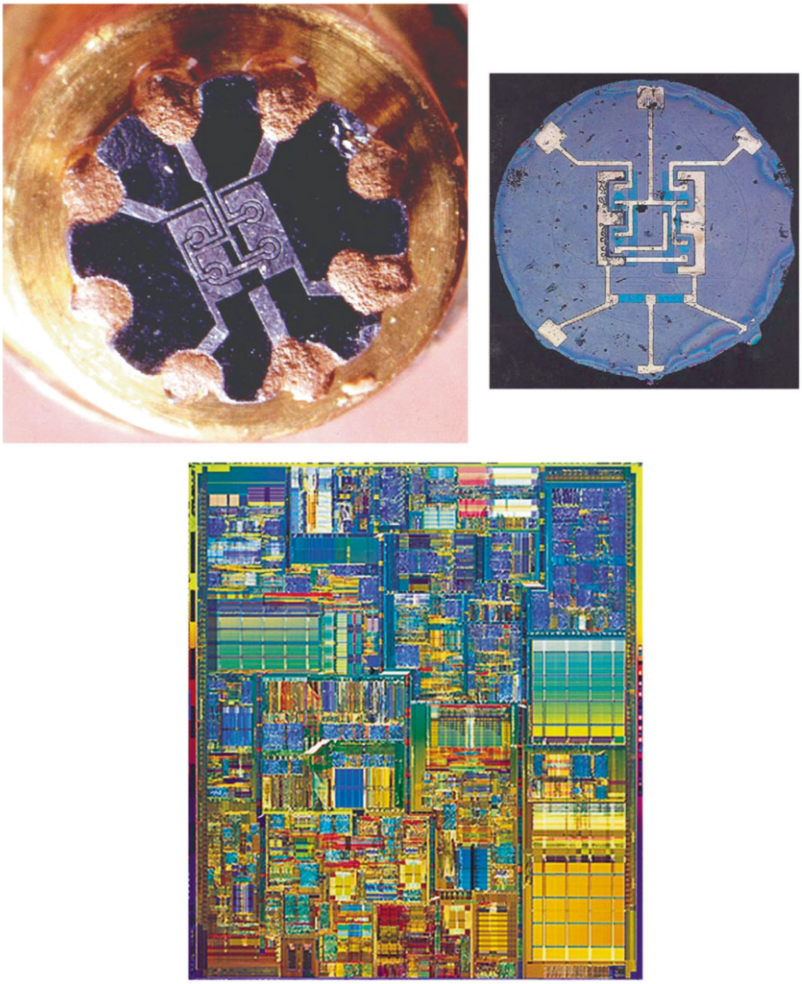


Fig. 2.11 First planar integrated circuit and one of modern variants of IC

it suffices to compare the first integrated circuit with the Pentium 4 processor, released in 2000, which harbors 4.2×10^7 transistors on the surface of 224 mm^2.

The final important step toward the creation of the modern semiconductor planar technology was the introduction of batch fabrication of integrated circuits, when a large number of identical circuits are made on the same substrate. This step was in fact quite natural given the large size and high quality of silicon wafers produced at that time (Fig. 2.12). Anticipating a little bit, we note that the size of available wafers grew rapidly—from wafers of ~25 mm in diameter in 1960 to 200 mm and more in the 1990s of the last century. At the same time the wafer area per one circuit went up from ~1 mm^2 in 1960 to ~100 mm^2 in the 1990s, with up to three million elements in each circuit.

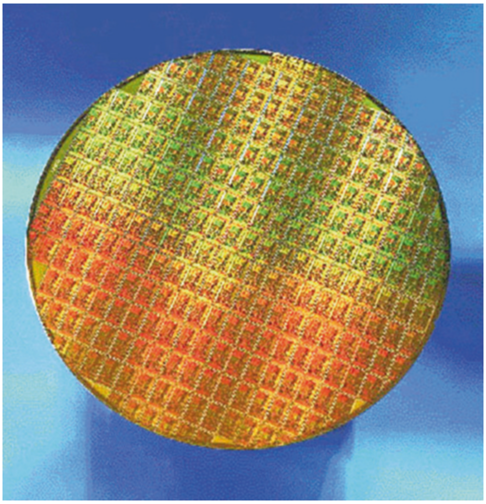


Fig. 2.12 Silicon wafer at the stage of producing integrated circuits

Today the semiconductor planar technology plays a central role in the production of electronic circuits (chips) that are used in a large number of devices—computers, control, and communication devices. It relies on a wide variety of specific processes that differ in their physicochemical nature and the instrumentation used. Typical for semiconductor technology are the extreme requirements imposed on the purity of raw materials, handling medium (water, auxiliary materials), and the atmosphere of the production facilities. In chemical practice, both in research and production, a substance is considered pure if the concentration of impurities does not exceed 0.001 %. The number of atoms in 1 cm^3 of the semiconductor is 10^{22}. Doping a semiconductor usually involves introducing 10^{16}–10^{19} dopant atoms per 1 cm^3, i.e., 0.0001–0.1 %. This means that the concentration of harmful impurities in silicon, which may affect its semiconducting properties, must be below 0.00001 %.

The planar technology is characterized by a number of other important features. However, since the main focus of this book is on the interaction and mutual influence of the computer technology and nanotechnology, we will confine consideration to the most important problem in this context—the lithographic process and the limits it imposes on miniaturization of electronic circuits.

2.3 Planar Semiconductor Technology: Universal Acceptance and Limitations

Planar technology (see Fig. 2.13) involves successive application on the surface of the silicon substrate of thin layers of material which serves to form the individual elements of the scheme, with subsequent processing of this layer.

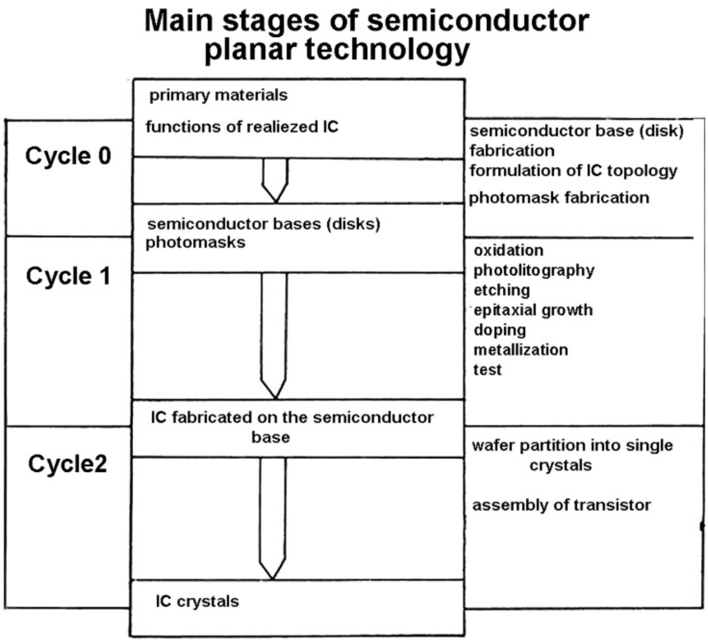


Fig. 2.13 Main stages of semiconductor planar technology

It is thus an approach utilizing the top-down principle. The essence of this principle, characteristic, for example, for the production of macroscopic parts of mechanical devices, is easy to illustrate using the manufacturing process of complex metal parts as an example. The process starts with the part blank which is successively subjected to turning, machining, boring required holes, cutting thread, etc. As a result of these successive operations, the part blank is turned into the final detail specified in the drawing. The planar technology uses various operations to create a layer of the required material. Insulating layers of silicon oxide are grown by controlled oxidation of the silicon substrate surface. Alternatively spraying, deposition from the liquid phase, etc., can be used to form the films as well. The main tool for creating individual elements of the chip is photolithography (Fig. 2.14). During the photolithographic process, photoresist is applied to a film of the material from which the elements of the integrated circuit are formed. Photoresist is a photosensitive compound which either decomposes upon light exposure (positive photoresist) or polymerizes, forming a solid film (negative photoresist). Photoresist is exposed through a photomask whose black and white pattern determines the shape and the location of details to be formed on the surface of the circuit. The photoresist film is subsequently "developed," that is, treated with solvent, which removes film areas unconverted by light. As a result, the photoresist film is transformed into a stable mask. Through its windows the material can be affected, e.g., by oxidation, doping, etc.

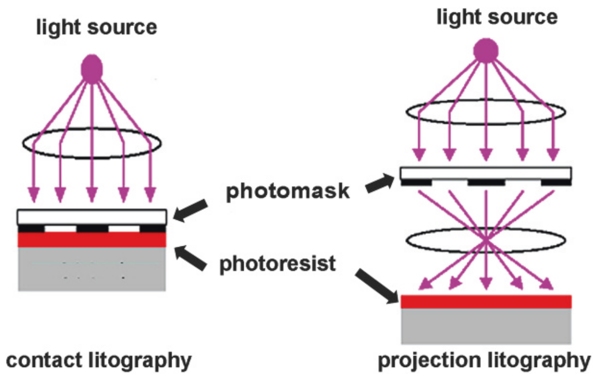


Fig. 2.14 Scheme of the photolithographic process

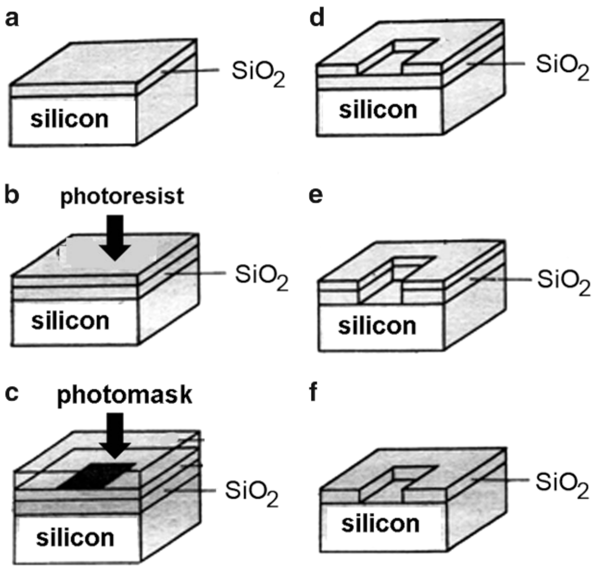


Fig. 2.15 Successive stages of the positive lithographic process (a–f)

As an example, consider the formation in a semiconductor of n-type zones corresponding to the source and drain of a planar transistor (Fig. 2.15). A negative photoresist is applied to the layer of silicon oxide grown on a silicon substrate. Non-transparent areas of the photomask used for illumination correspond to the n-type areas to be formed. As a result of illumination of the photoresist its entire surface polymerizes except for those areas being created, from which unreacted resist is cleaned up with a suitable solvent (benzene, toluene). Then, through the windows in the polymerized film of the resist, silicon oxide is removed with hydrofluoric acid, after which the polymerized resist is removed by a new solvent. In this fashion windows are formed in a layer of silicon oxide on the surface of the integrated circuit that are used for doping silicon by phosphorus or antimony. The process of doping involves diffusion: alloying additive is applied to the surface and

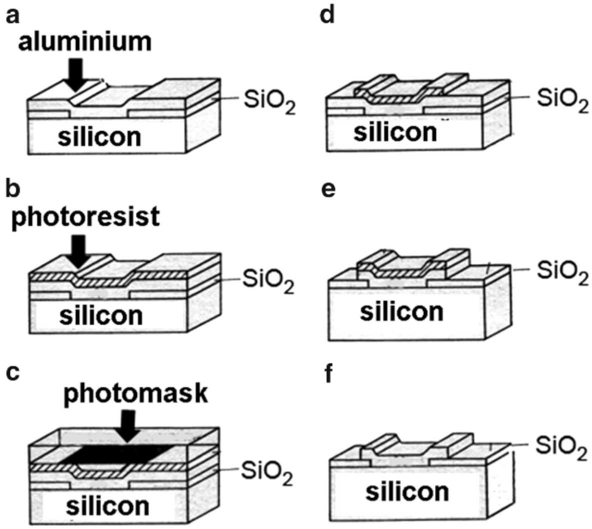


Fig. 2.16 Successive stages of the negative lithographic process (a–f)

then the chip is heated. Alternatively ion implantation by direct action of dopant ions can be employed. Figure 2.16 illustrates the positive lithographic process in which a layer of aluminum deposited on the surface of the chip is removed from all its elements except for the gate area of the field-effect transistor.

Various resists are known which differ in terms of their light sensitivity and resolution. Often a positive photoresist PMMA (polymethylmethacrylate, Fig. 2.17) is used. As a negative resist, COP (a polymer of glycidyl methacrylate and ethyl acrylate) is used.

Photolithographic technology allows for the creation of complex semiconductor circuits, but at the same time it represents the limiting factor of the planar semiconductor technology. Due to its wave nature light is diffracted by the photomask elements (Fig. 2.18). The maximal achievable resolution of the exposed pattern is of the same order as the wavelength of the light used for illumination. As is known, the wavelength of visible light is in the range from 0.38 μm (violet region of the spectrum) to 0.76 μm (red area). This determines the minimum line width in a semiconductor structure which can be obtained by optical lithography.

Today's scientific and engineering projects are aimed at improving the key step in the production of integrated circuits—lithography—which will determine the physical limits of semiconductor technology in the foreseeable future. Experts note that because of these restrictions lithography may exhaust its possibilities already in the beginning of our century.

The development of the lithographic technology since its inception in the early 1970s was directed at reducing the wavelength of the light used. This allowed to reduce the size of the elements of the integrated circuit. Since the mid-1980s ultraviolet laser radiation is used in photolithography. In order to apply the pattern of the circuit to the plate, computer-controlled machines (steppers) are employed. Configuration of "window" is determined by appropriate masks, and the resulting

Fig. 2.17 Positive (**a**) and negative (**b**) photoresists

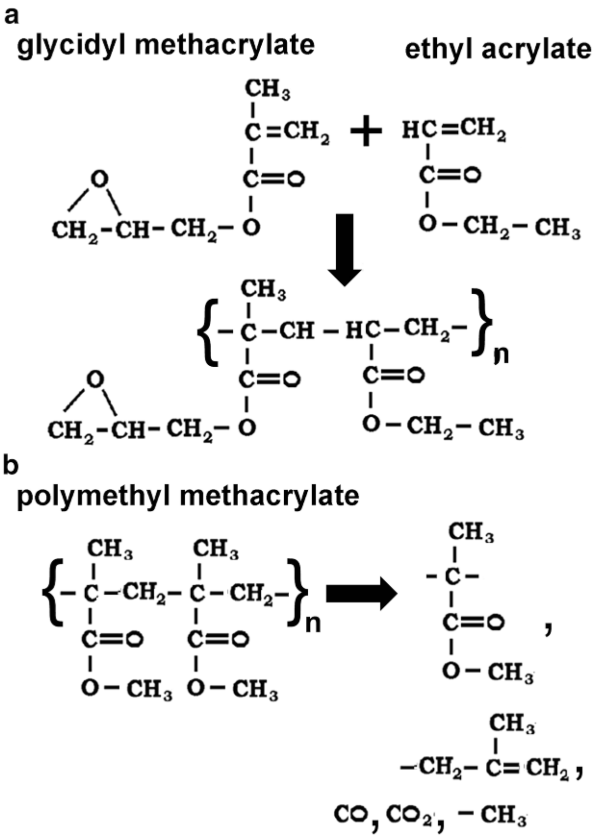


Fig. 2.18 Limitations of photolithography

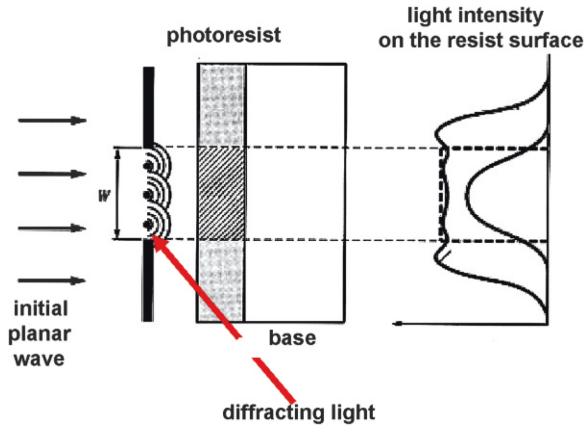


image is focused by a special lens system which reduces the given mask pattern to a microscopic size. A modern photolithography machine handles several tens of 8-in. semiconductor wafers per hour.

Currently, most chips are made using ultraviolet radiation with a wavelength of 0.248 μm. For some circuits, a lithographic technology with the wavelength of 0.193 μm has been developed. However, beyond 0.2 μm, serious problems put under question further progress in photolithography. For example, at a wavelength of less than 0.2 μm too much light is absorbed by the photosensitive layer, complicating and slowing down the process of transferring the pattern of the circuit template. Such problems motivate investigators and manufacturers to seek alternatives to conventional lithographic technology. For example, the possibility of replacing ultraviolet rays by X-rays has been under investigation in US scientific laboratories for more than two decades.

One technology, called EUV (extreme ultraviolet) and supported by several well-known companies, aims to improve the process of lithography in chip manufacturing.

As already noted, modern equipment for printing circuits on silicon substrates based on deep ultraviolet radiation (deep ultraviolet, DUV) uses light sources with a wavelength of 248 nm. It is assumed that the wavelength of EUV radiation can be as short as 13 nm, i.e., approximately 20 times shorter. The transition from the DUV to the EUV lithography provides for more than tenfold decrease of the wavelength, making it comparable to the size of just a few tens of atoms.

Nevertheless, in addition to purely physical problems, there are other factors in the manufacturing process of circuits limiting miniaturization and the degree of integration of transistors. Generally speaking, the properties of the devices created on the same silicon wafer, as well as on different wafers, are not identical. Deviations can occur at each stage of production. The nature of the possible differences between the produced circuits and the frequency of occurrence of completely defective devices may hamper further miniaturization of integrated circuit elements. Note that miniaturization affects not only the length and the width of the circuit but also the thickness of the crystal on which transistors and connections are implemented through a series of levels. In modern chips, there may be four or five such levels. Reducing the size of transistors and increasing their density on the crystal brings about an increase in the number of levels. However, the more layers exist in the circuit, the more thorough control of the production process is required, since each of the levels will be affected by the levels underneath it. The cost of improving control and creating connections between multiple layers may deter the increase in the number of layers.

Among other things, the increasing complexity of integrated circuits necessitates further improvement of production conditions, to which unprecedented requirements are already posed. A more precise mechanical control over the positioning of the original silicon wafer is necessary. Sterile rooms (so-called clean room) in which chips are manufactured should become even cleaner to exclude the penetration of tiny dust particles that can destroy a complex circuit.

Taken together, all this not only demonstrates the need to improve the planar semiconductor technology but also motivates the quest for fundamentally new approaches to building computing devices. One of the most promising ones is the transition from semiconductor to molecular components.

Springer

http://www.springer.com/978-3-211-99698-0

Molecular Computing
Origins and Promises
Rambidi, N.G.
2014, XIII, 217 p. 143 illus., 34 illus. in color., Hardcover
ISBN: 978-3-211-99698-0