The Semiconductor Story

1: The new crystal triode

by K. J. Dean, M.Sc., Ph.D., and G. White†, M.Phil., B.Sc.*

The paper which first announced the discovery of the transistor appeared in the *Physical Review* in July 1948. To commemorate the 25th anniversary of this event, *Wireless World* is publishing a series of four articles presenting a critical survey of the semiconductor industry, past and present, from the U.K. point of view. Part 1 describes the early development of germanium diodes and transistors, while parts 2 and 3 describe respectively the exploitation of the transistor and the integrated circuit to the present day. The final part discusses some of the problems, both technical and commercial, which have faced the industry in recent years. The roles of careful research, happy chance, technical skill and industrial pressure make a fascinating story of our times.

The new crystal triode, as the transistor was first called, seemed in 1948 to be poor competition for the Goliath sized valve manufacturing industry. But a veritable David it turned out to be! *Wireless World* reported the discovery in an article in October 1948, entitled "The Amplifying Crystal". How many people reading that report then realized its implications for the future? The transistor was the end result of research which started 140 years ago in 1833 with Michael Faraday. He noted that while most conductors have a positive temperature coefficient of resistance, a substance called silver sulphide had a negative coefficient. Thus a substance later to be classed as a semiconductor was identified. Rectification, photoconductivity and photo-e.m.f. effects were all observed before 1900. Theoretical work on semiconductors after Faraday's original discovery gathered momentum, so that, by the early 1930s, quantum mechanics was applied to the theory of conduction. Energy band diagrams, electrons and holes then started to be discussed. The stage was set for the discovery in America by J. Bardeen and W. H. Brattain of the transistor—a semiconductor triode. This was the first three-terminal semiconductor device which could amplify, and that was only 25 years ago. Now the impact of the transistor is universal, it has applications ranging from aviation and broadcasting to washing machines and Xerography.

Cat's whiskers

Semiconductor crystals were used in the early days of radio communications, the crystal rectifier being used as the detector in radio receivers. A typical detector was made by soldering or clamping a minute piece of the crystal in a small brass cup and the point contact made with a flexible wire called the cat's whisker, which was held in light contact with the crystal. The discovery of the thermionic triode by de Forest in 1907, and its subsequent developments, made the crystal rectifier obsolete in radio receivers. However, the point contact crystal could not be replaced for detecting and monitoring u.h.f. power. At the other end of the scale, at low frequencies, the copper oxide rectifier and selenium rectifier have been commercially successful, but they are however not point contact rectifiers. The rectification property of these is obtained by the contact of a thin film of semiconductor with the metal on which it is deposited. They are therefore termed contact rectifiers.

Wartime research

The second World War, like all military ventures, provided the cash to oil the wheels of research, so important at times of national emergency. It saw the development of radar, which gave a great impetus to u.h.f. crystal rectifier design. Research was concentrated on using silicon, germanium and boron. Boron prepared with selected impurities, i.e. "doped", showed sufficient conductivity to be of interest, but its typical characteristic curve was S shaped and symmetrical about the origin, thus the project was then dropped. Silicon showed great promise, being used for most of the commercially available devices. At this time the importance of starting with extremely pure silicon was appreciated. The "red-dot" crystal diode developed by the General Electric Company, for example, was derived from silicon crystals prepared from melts made from highly purified silicon powder, to which was added a fraction of a per cent of aluminium and beryllium. The resulting crystal could dissipate relatively large amounts of power without appreciably impairing its performance as a mixer. These were therefore known as "high-burnout" crystals.

The method of adjusting the cat's whisker at this time is interesting to note. The contact pressure was increased until a predetermined characteristic was obtained, and the cartridge was then tapped with a light mallet. Careful tapping caused the forward resistance to drop and the reverse resistance to rise. The cartridge was then impregnated with wax to provide mechanical stability and to make it impervious to water. Further work in 1943 led to high purity silicon, doped with only 0.001% boron, which produced an extremely good device and made prolonged tapping unnecessary. The small amount of the impurity needed indicates how material technology had to keep pace with the demands of the semiconductor device manufacturer. At this time, work on germanium led to the high-inverse voltage rectifier; so called because it could withstand up to 100V applied in the reverse direction. The doping agent used was tin, although it was found that similar effects could be obtained with some other elements. Germanium, however, could not compete with silicon above 30MHz. These methods of preparing the germanium crystal and polishing its surface were to be used later in the manufacture of the first transistor.

In 1946 H. Q. North showed that the point-contact used in these devices could be welded to the crystal surface, by passing a high density current (in the order of 10^7 amps/sq. in) for a short time through the contact point. Although this did not improve their performance, little was lost either. This technique too was later to be of value in three-terminal point contact devices.

Post war development

After World War II the immediate problems of survival gave place to the interests of commercial enterprise, and researchers were able to return to more general semiconductor problems, although under industrial patronage. Silicon and germanium were chosen for the research effort, because they are simpler to understand than most other semiconductors. A lot of expertise on these materials had been accumulated during the war, particularly in America. Fig. 1 shows the structure of silicon or germanium crystals. Each atom has four neighbours, all

---

*South East London Technical College.
†Twickenham College of Technology.

at the same distance from it, and all at equal distance from each other. Each atom and one of its neighbours is attached by an electron pair bond, which consists of sharing two electrons to form a stable bond. Each atom has four electrons available to form bonds (valence electrons), therefore the conditions are exactly right for the diamond structure of Fig. 1.

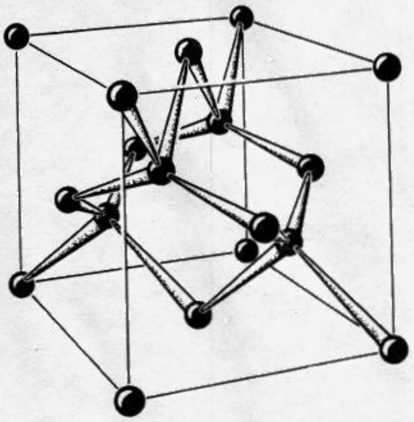


Fig. 1. The crystal structure of germanium and silicon.

The electronic properties are also dependent upon the electrons present in the bonding. By introducing impurities into the crystal the bonding can be modified. Therefore, the electronic properties can be tailored as required by the controlled addition of impurities. The unoccupied bonds on the extreme edge of a perfect crystal cannot be used by internal atoms, but they are capable of accepting electrons. These are called acceptor or surface states. Crystal defects and absorbed foreign atoms will have similar effects and also create surface states. It was the thorough investigation of these states that led to the somewhat accidental discovery of the transistor effect. It is strange that surface states are now something to be avoided in transistor manufacture, because they would provide a low impedance path to current flow that is controlled inside the material.

Amplification using semiconductors was first achieved by using the negative resistance characteristic of thermistors. As the current through the thermistor increased, the heat generated caused a reduction in the resistance, and hence a drop in the voltage. The frequency of operation is limited by the temperature which has to follow the current changes. However, by making the physical dimensions small and the thermal conductivities high, oscillations of up to 100kHz have been produced. Bell Telephone Laboratories' aim after the war was to produce a purely electronic, rather than thermal, semiconductor amplifier. The work was initiated by W. Shockley who directed work on investigating the modulation of the conductance of a thin film of semiconductor. The conductance was controlled by an electric field applied by an electrode insulated from the film. It was hoped that the conductance would be modified by changes in the surface states caused by the applied field. The experiment gave disappointing results, since only about 10% of the expected change in conductance occurred. The effect was explained by J. Bardeen who in 1947 proposed a double layer at the surface, formed by the charge in the surface states and the induced space charge. Further research was carried out to measure the characteristics of the surface states.

The transistor discovered

The effect of having the crystal surface immersed in a liquid was studied. The characteristics of a high-inverse voltage germanium rectifier with a field applied by an electrolyte were investigated by J. Bardeen and W. H. Brattain. They proposed that a portion of the current was being carried by holes flowing near the surface. When the electrolyte was replaced with a metal object, transistor action was discovered. The discovery was first published as a short letter to the editor of the *Physical Review* journal in July 1948. This marked the beginning of the transistor era. A more detailed paper was published in the following year.

The transistor is a semiconductor triode amplifier. The prefix "trans" designates the translational property of the device, while the root "istor" classifies it as a circuit element in the same general family with resistor, varistor, and thermistor. The transistor was commercially made in a similar form to the point contact diode, except for a second cat's whisker mounted very close to the first. The device is shown schematically in Fig. 2. A germanium ingot was prepared in the same manner as that used for the high inverse voltage diodes, and then a slice of this ingot was ground flat on both sides. The slice was copper-plated and tinned on one side, and diced into small squares with a diamond wheel. One of these squares was then sweated onto the brass base plug and the germanium surface treated. The unit was force fitted into a cylindrical cartridge, which had been shaped to accept the contact assembly. The contacts consisted of two 0.005in phosphor bronze wires, which had been bevelled and polished.

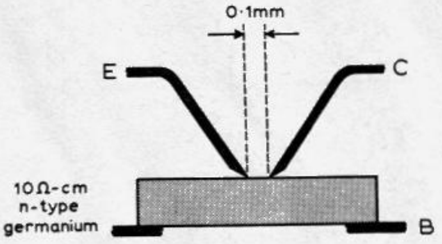


Fig. 2. Schematic of the point contact transistor.

The characteristics of the thermionic diode and the semiconductor diode are fairly similar, and methods of adding a "grid" to control the current in the forward direction as had been achieved with the triode, were looked at. The transistor, however, is not operated in this quadrant, because the output is reverse biased in the high resistance direction. The current is enhanced and controlled by the forward biased emitter contact. This device was designated the type A transistor to distinguish it from possible future varieties. The transistor effect is the injection of holes into the n-type material by the emitter, which are collected as an increment of the collector current. The common terminal called the base electrode is physically the base of the crystal. Devices which operate on different principles, such as the field effect, have since been called transistors. Therefore, transistor electronics is used generally to describe the art of controlling electron movements in a solid, hence is sometimes called solid state electronics. One of the first point contact transistors to be manufactured in the United Kingdom is illustrated. The patent numbers show the advantage of a strong development facility, by using experience gained in the construction of point contact diodes to help in the manufacture of transistors. Patent number 591092, which was applied for in 1945, describes a method for holding the contact in place after construction. This is achieved by filling the cartridge with a wax-like substance which will harden on heating. The other patent number, 592659, was applied for in 1941, and deals with the preparation of the crystal and the subsequent treatment of its surface. The germanium had to have a spectroscopic purity of 99.95% for good results.

The G.E.C. crystal triode type GET 1, one of the earliest point contact transistors to be made in the U.K. The reverse of the packet, shown here with the transistor, carried a warning "To prevent permanent damage to the triode, it is recommended that whenever possible d.c. limiter resistors be placed in series with both emitter and collector . . . Great care should always be taken to connect supplies of the correct polarity to the electrodes."

Transistor amplifiers

The journal *Audio Engineering* published an article in August 1948 entitled "Experimental Germanium Crystal Amplifier", only one month after Bardeen and Brattain's original letter. This described how to construct a germanium crystal amplifier—such was the rate of progress even in 1948. The article highlights the similarity between point contact diodes and the type A transistor because the construction starts with two diodes. They are dismantled and the crystal used, with the two whiskers carefully adjusted on the surface. Difficulty was experienced in finding active spots, due to the relatively impure crystals being used at that time. Manufacturers were aware of the need for high quality germanium. In 1946 the first extraction plant in the United Kingdom was built at Brimsdown for Johnson Matthey for the bulk production of germanium and other semiconductor materials.

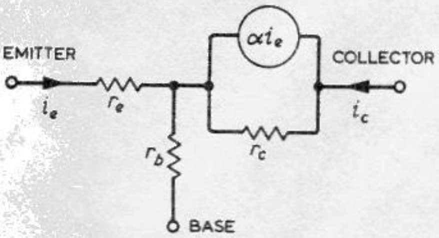


Fig. 3. Equivalent Tee circuit of a transistor.

The type A transistor can be represented by the equivalent circuit shown in Fig. 3, with the following average values for its parameters:

emitter resistance $r_e = 240\Omega$

base resistance $r_b = 290\Omega$

collector resistance $r_c = 19{,}000\Omega$

amplification factor $\alpha = 1.8$

Unfortunately the active area of the device is very small and hence the collector dissipation is only about 0.2W, although a power gain of 17dB with a power output of 5mW was achieved. The small size of the device, however, gives it a wide frequency response, with an upper limit of approximately 10MHz. It was soon noted that the transistor could be greatly improved by passing large reverse currents through the collector point. This technique, called forming, resulted in amplification factors as high as 5. This process was explained by the formation of a p–n hook at the collector which reduced the height of the potential energy hill at the collector, so allowing a considerable increase in the number of electrons diffusing from the collector into the floating p region.

The movement of holes was thought to be mainly confined to the surface region but in 1949 J. N. Shive proved that the flow of charges could be through the bulk of the material. This was shown by constructing the double surface transistor, which was produced with germanium in the shape of an acutely tapered wedge, the two contacts being opposite each other near the thin edge. This transistor was developed into the coaxial transistor which was much easier to manufacture. Here the germanium was cut into a pill shaped cylindrical wafer with a dimple ground into the centre of both sides, so that the thickness of the centre was only a few thousandths of an inch. The emitter and collector contacts then bear on opposite sides of the semiconductor in the dimples, and are arranged coaxially to fit into a cartridge. This method of construction avoided the problem of placing two spring contacts within a few thousandths of an inch of one another. The components used were similar to the parts used in the manufacture of point-contact rectifiers.

Junction transistors

In 1949 W. Shockley proposed that transistor action could be achieved with p-n junctions within a single crystal, thus breaking away completely from the surface effects of point contact devices. The device was therefore called a junction transistor. In principle it consisted of a bar of single crystal n-type germanium, for an n-p-n device. In the centre of the bar was formed a thin layer of p type germanium as part of the single crystal. Ohmic non-rectifying contacts were attached to each of the three regions, the outer two being the collector and emitter and the centre the base. The method of operation is essentially the same as the point contact, hence the electrodes have the same names, although the base is now in the middle. The equivalent circuit chosen for comparison is the same one as used for the point contact i.e. Fig. 3; the new values for the parameters are:

$$r_e = 25\Omega$$

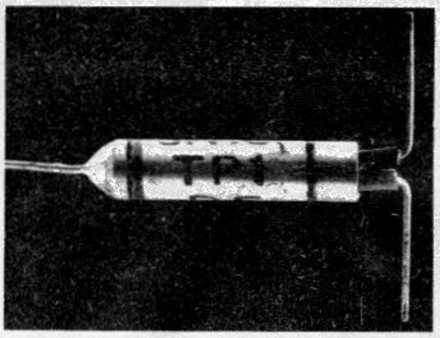
$$r_b = 250\Omega$$

$$r_c = 5\times10^6\Omega$$

$$\alpha = 0.95$$

The amplification factor α for junction transistors is less than unity, hence the amplification in common-base operation is due to the difference in impedance levels. A junction transistor was developed with a p-n hook collector, which acted similarly to the point contact transistor as far as the gain was concerned. This was achieved by a four-layer p-n-p-n device, but the transistor had a poor high-frequency response. Little further work was carried out, even though high amplification factors were obtained.

The first junction transistors were a great improvement over the point contact devices. Power gains of 40dB, with class A operation of 49% efficiency were achieved against 23dB gain and an efficiency of 30% for point contact transistors. The higher power gain is due to the increase in the output impedance, and the almost ideal characteristics show that the junction transistor can operate close to the 50% maximum for a class A amplifier. Junction transistors will operate with extremely low input power of around 0.6μW. This is about one ten-thousandth of the power required to operate the point contact transistor, or one millionth of the power to heat the cathode of a typical thermionic valve. Unfortunately the frequency of operation at that time was limited to about 1MHz. This was due to the time taken for the charge carriers to diffuse across the base. The equivalent effect in thermionic valves is the transit time, that is, the time taken by the electrons to travel from the cathode to the anode. The type of case used by S.T.C. for an early junction transistor is shown in the photograph. Although the TP1 device shown was a point contact transistor, it was made at the same time, and externally looks identical to the TS1 junction transistor.



The S.T.C. point contact transistor TP1 appeared about the same time as the G.E.C. GET 1. It was soon withdrawn and replaced by the TS 1, a junction transistor.

Several methods have been used to improve the high-frequency response of junction transistors. The most obvious answer is to reduce the base width; this is limited, however, by the problem of punch through. A second contact added to the base by Wallace *et al* in 1952 effectively reduced the base area and the base resistance. This increased the cut-off frequency to about 50MHz. Further improvements were realized by advances in material technology, in particular by the diffusion process which started in 1952, and by the production of extremely pure silicon. The purification was achieved by zone refining. This process is based upon the relatively high rate of diffusion of impurities in the molten zone of a crystal, compared with the much slower rate in the solidification zone. The raw single crystal is passed slowly through a localized radio-frequency heating coil. The crystal within the coil is in the molten state, and on passing through the coil re-solidifies into a single crystal again. The impurities tend to remain in the molten region and therefore are swept to the end of the crystal. The process is repeated several times. The end with the impurities is discarded and the concentration of impurities in the main section can be reduced to about 10^{17} atoms/cu.m.

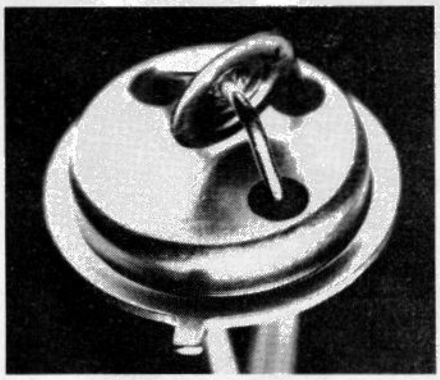
Field effect

The field effect transistor experiments that failed were the beginning for the point contact and junction transistors. In 1952 W. Shockley proposed a unipolar field effect transistor which overcame the earlier problems of surface states. The point contact and junction transistors are called bipolar because charge carriers of both signs are involved. In the field effect the controlled conductance between input and output terminals results from changes in the number of carriers of one type, hence the name unipolar. The field effect transistor has several advantages, the most important being the high input impedance. The input is a reverse biased p-n junction, and the depletion layers created control the conductance through the channel. The difference in operation is reflected in the names for the electrodes, the emitter and collector being called the source and drain respectively. The controlling electrode is now called the gate instead of the base. It was not until fairly recently that the technology needed to be able to mass produce these devices has been developed. In the meantime the junction transistor has built up a commanding lead.

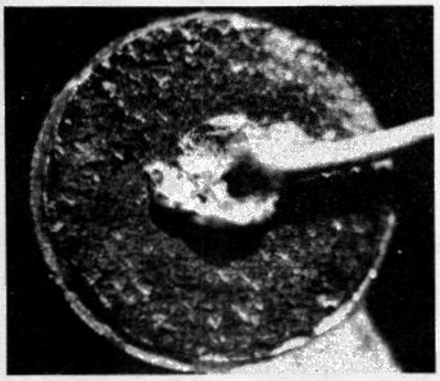
Circuit design

Early work on transistor circuit design tended to start with a well tried thermionic valve circuit, and then modify it for use with transistors, even though the parameters are radically different. The grounded cathode triode is a voltage amplifying device with a high input impedance and a relatively low output impedance. Conversely the grounded base transistor is a current amplifier with a low input impedance and a relatively high output impedance. The early papers on transistor circuit design referred to the transistor's characteristics as peculiar, because they were different to those of a valve. On looking further at the parameters, it was noted that, if the roles of current and voltage were changed over, the devices were similar enough for quantitative designs starting from the valve circuits. This background led to the circuit performance of transistors being less than they might have been, until designers began to take account of the transistor's peculiarities and use them to advantage. One of the major advantages which would be unheard of with valves is the use of complementary circuitry, allowed by having n-p-n and p-n-p transistors.

The small size and ruggedness of transistors opened new fields and their small power requirements meant that the components used with them could be miniaturized also. The type A transistor of 1949 occupied one-fiftieth of a cubic inch, with a collector voltage of 30V. In 1952 the junction transistor could be fitted into one five-hundredth of a cubic inch with a collector voltage of 2V. Bell Telephone Laboratories studied the problem of manufacturing complete circuit packages under an American Signal Corps contract in 1952. At that time the package of a laboratory circuit model required about one-tenth the space and power of an equivalent package built with thermionic valves. The importance of designing sub-sections of a system, which would be used in quantity, and manufacturing them as packages was realized from the beginning of the transistor's development, and has been a goal ever since.



A modern germanium alloy junction transistor still in production at Newmarket Transistors. The emitter lead is in the foreground and the base lead at the right connects to a metal disc in which the semiconductor pellet is held.



A photomicrograph of an early medium power germanium alloy junction transistor. The pellet of impurity and the emitter lead connected to it are clearly shown in the centre of the picture.

The general manufacture of transistors began in 1952, after Bell Telephone Labs. held a symposium, where they offered know-how to all who wanted it for the price of an admission ticket ($25,000). The era of the practical transistor had now begun. Photographs show the construction of an early alloy junction and the progress achieved since then by comparison with a modern alloy junction transistor. The successive developments to improve the parameters and to find transistor structures, which lend themselves to easier manufacture are related in part 2 "The search for the best transistor". The originators of transistor electronics, J. Bardeen, W. H. Brattain and W. Shockley were awarded the Nobel prize for physics in 1956 in recognition of their work in the theory of semiconductors, when it was beginning to be recognized that they had not just invented the transistor, but had laid the foundations of the world-wide multi-million pound microelectronics industry.

(To be continued)

Bibliography

1. "The amplifying crystal", *Wireless World*, October 1948.
2. Bardeen, J., "Surface states and rectification at a metal semiconductor contact", *Phys. Rev.*, 71, pp. 717–727, 15 May 1947.
3. Bardeen, J., and Brattain, W. H., "The transistor, a semiconductor triode", *Phys. Rev.*, 74, pp. 230–231, 15 July 1948.
4. Shockley, W., "The theory of p-n junctions in semiconductors and p-n junction transistors", *Bell. Sys. Tech. Journ.*, 28, pp. 435–489, July 1949.
5. Shockley, W., "Electrons and holes in semiconductors", Van Nostrand, 1950.
6. Morton, J. A., "Present status of transistor development", *Bell Sys. Tech. Journ.*, 31, pp. 411–422, May 1952.
7. North, H. Q., "Properties of welded germanium rectifiers", *Journal of Applied Physics*, pp. 912–923, Nov. 1946.

Sixty Years Ago

An uneasiness in the Marconi Company when share prices fell considerably is reflected in a statement issued by the company and reproduced in the January 1913 issue of *The Marconigraph*. It referred to opinions which had been expressed suggesting that "continuous waves would in the future supersede the spark system". The announcement from the secretary of the comapny continued "As these statements and opinions are liable to mislead shareholders and cause them some uneasiness, I am instructed to inform you that Mr. Marconi himself tested continuous wave systems many years ago, and experimented with them during the greater part of 1907 at the Poldhu station. As a result of these experiments he learned the advantages and disadvantages pertaining to continuous waves, and eventually arrived at a compromise between the continuous wave and spark systems, combining the best points of both. This resulted in material changes in his system for long distance work, and new and important improvements were patented by him in 1907, which are mainly responsible for the progress since made in long-distance wireless telegraphy. These inventions, which materially modify the spark system, seem to be surprisingly little known, notwithstanding the lectures delivered by Mr. Marconi ....when he made statements relating to the use he was making of continuous waves, semi-continuous waves and the elimination of the spark."

Corrections

L. Nelson Jones, author of the article "I.C. Peak Programme Meter" in the November 1972 issue, has informed us of an error in the specification of the meter. The scale marking division seven represents a level of +12dBm (not 14dBm) with a peak input voltage of 4.38V. The undefined f.s.d. reading usually corresponds to around 5.37V peak. This calibration fault is easily corrected by changing the value of R_{14} to 100kΩ, and Key Electronics, suppliers of the kit, are sending all those who have kits, a replacement resistor together with a copy of the amended handbook.

We regret an error exists in the circuit diagram (Fig. 1) of the **"Mobile/Portable Power Unit for H.F. Transceiver"** published in our December issue. The conductor between the base terminal of transistor Tr_2 and ground should be omitted otherwise the catastrophic failure of this device will occur.

The Semiconductor Story

2: Search for the best transistor: continuing a four part series of articles commemorating the 25th anniversary of the transistor

*by K. J. Dean**, *M.Sc.,Ph.D., and G. White*†, *M.Phil.,B.Sc.*

At the start of the 1950s the transistor was a novelty. Industry needed to be convinced of its advantages over valves and electromechanical devices such as relays and magnetic amplifiers. Besides, there were a number of types being developed—which was the best? Even the textbooks of the period hedged their bets, taking as much space over point contacts as over junction transistors. But the electronics industry, at least, was just beginning to take notice. In 1952 the Post Office Research Station at Dollis Hill had demonstrated the first line amplifier to be made in the U.K. which used junction transistors, while a year later in America, Texas Instruments produced their first pocket transistor radio.

1953 was an important year for the U.K. semiconductor industry. One might almost say that was its birth, for in that year a number of companies set up manufacturing plants, among them G.E.C., Mullard, Ferranti and Pye, who were not then in the Philips group. One of the problems at that time was that the available germanium transistors did not have worthwhile gain at radio frequencies. Naturally, therefore, one of the first commercial applications that they chose to exploit was that of transistor amplifiers for hearing aids. The Post Office was the authority for National Health hearing aids and under its guidance Mullard developed the OC56 and OC57 junction transistors specifically for this market. At the same time, Pye at Cambridge had interested Acousticon Ltd, manufacturers of valve-operated hearing aids, in transistors and the first 300 were delivered at the end of 1955. Some of these early devices were packaged in glass cases which were filled with silicone grease and were then painted to prevent the photoelectric effect (amplified by the transistor) making the other current changes due to transistor action. Many an engineer carefully scratched the paint away to use them as sensitive photocells until the manufacturers foiled this dodge by using metal cans. Some of the first metal cases were sealed with solder, leading to examples of flux contamination. The Post Office was not satisfied with these types of encapsulation and insisted on hermetic sealing.

So difficult was the technology of junction devices to master that one manufacturer in those early days recorded that the yield in the first week of production was one device and another calculated that his first working transistor represented an investment of £1 million.

One seldom stops to think why the U.K. semiconductor industry developed as it did. Where did the money come from? Who made the decisions that got it all started? Many companies owed their place in transistor research to the encouragement of C.V.D. (Commercial Valve Development!) This government committee, on which the services, the Post Office and our national research establishments were represented, placed contracts for the development of transistors. It is always popular to blame government for wrong decisions or for no decisions at all, but without C.V.D. help few U.K. companies would have got started. One exception was Mullard, owned by the Dutch Philips Group, whose research was funded from the profits of selling valves. In fact their early transistors used valve nomenclature: A for diodes, B for double diodes and C for triodes. The first symbol of the type number was reserved for the heater voltage, zero for transistors of course. So the OC70 was clearly a triode with no heater.

Difficulties with germanium

The first transistors were germanium devices but for a long time the material which would eventually be best was in doubt. Supplies of germanium were limited as there were only three known ores. Two sources were in Zaire (then Belgian Congo) not a particularly stable part of the world; a third ore, germanite, came originally from South Africa, but the mines were exhausted there so that its chief source was from ores imported into Germany before World War I. In addition certain coals contain germanium and at that time the principal supplier in the U.K. was Johnson Matthey who indicated that their main source was from flue dust. Hence, the price of pure germanium was high—about £100 per lb. Meanwhile in Japan the Tokyo Gas Company was extracting germanium from waste coal-gas liquid—one of the first signs of competition from the Far East. It was estimated that one ton of germanium would make 200 million transistors and that in a few years 40 tons per annum would be needed for the world market, against the current production of three tons per annum, including the germanium needed for other purposes. Something had to be done.

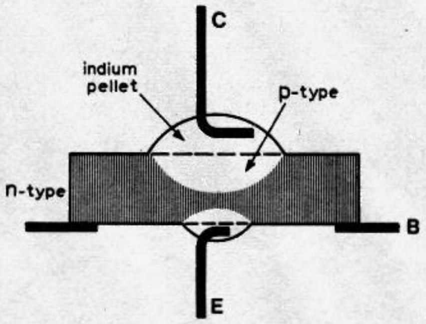


Fig. 1. Slab of n-type germanium with two indium-doped pellets alloyed to it so that it will be modified to p-type immediately below them after heating. The resulting alloy junction transistor was illustrated by a photomicrograph in Part 1 of this series.

Silicon was the obvious contender. Like germanium it is a group IV element; also, after oxygen it is the most common element in the earth's crust, but its melting point is 1420°C compared with 937°C for germanium. The purification of germanium requires a heating and cooling cycle of seven hours, one hour of which was at 1050°C in an atmosphere of pure dried hydrogen. The temperatures for silicon are correspondingly higher. Large quantities of expensive argon are used, which had to be reclaimed, and there were difficulties with phosphorus and boron impurities. Also the quartz (that is, silica) of the crucibles used tended to dissolve in the silicon. As late as 1955, S.T.C. (Standard Telephones and Cables) reported that their own attempts to purify silicon to the extremely high standard of purity required had not been successful. "No further work was done," the report adds, "due to the loss of the man doing it." Nowadays a large proportion of manufacturers are content to buy-in purified semiconductor material in slices for them to process.

Successes with silicon

Texas Instruments were first in the field with silicon transistors in 1952 and had a virtual monopoly for three years. At first the current gain was low and the frequency response was poor due to the lower mobility of charges compared with germanium.

---

*South East London Technical College.
†Twickenham College of Technology.

There were difficulties in controlling the technology, but leakage currents, always a difficulty with germanium, were much less. Ferranti, which had not until now been in semiconductors, decided to work solely with silicon (except for a small production of germanium tunnel diodes) on entering the field.

Difficulties with materials were by no means the only problems: there were insistent demands for higher frequency operation and higher power also. Receivers at that time were even being designed with valve "front ends" and transistor audio stages operating earphones. In 1954 S.T.C. had joined the semiconductor club, much of the work being done in germanium at the Brimar Valve Company's Engineering Division at Footscray, the basic research going on at Enfield and Ilminster. Their first junction device was the 3X/300N, later renamed TS1. It had a rating of 50mW while Philips cautiously rated their transistors at only 6mW, although after 15 months of life tests they were upgraded to 25mW. Pye moved their semiconductor plant to Newmarket primarily to develop a solid-state radio which was marketed by Pam in 1956. Meanwhile in Japan, Sony had started manufacturing transistors in 1953. A year later, they produced their first transistor radio and so started a virtual monopoly of short-wave and f.m. transistorized receivers, which was to last a decade. At this time, the best that the U.K. could offer was the V6/R2 of Newmarket Transistors and OC44 of Mullard, both of which had $f_T = 6$MHz.

New types of transistors

The first junction transistors had grown junctions, produced by overdoping, in which the predominant impurity of the melt was interchanged at regular intervals as the crystal was drawn from it. The method was unsuitable for quantity production. The characteristics of these transistors left much to be desired—with light doping at the start and heavy doping with correspondingly lower resistivities at the end of the pull. Consequently, the alternative method of alloying which had been known since 1948 was the one which was principally developed and which resulted in most of the devices described earlier. In this process, small pellets of impurity material are fused to one side of the germanium slice and somewhat larger ones to the other side to form emitters and collectors respectively. For p-n-p transistors indium was used and lead-antimony pellets for n-p-n types. Subsequently, the slice was cut up into chips. It was an adaptation of this process which seemed to offer the best solution to higher frequency operation. This was the alloy diffused process developed simultaneously in Holland and in the U.K. (by Julian Beale) by Mullard.

The alloy for one of the pellets was a mixture of two impurities. There was a fast diffusing n-type impurity to define the base, with a slower diffusing p-type material. Hence, on heating, the first diffuser goes ahead of the alloy front. This process produced a graded base in which carriers crossed the base region more quickly than in the simple alloy types. Furthermore, the process lent itself to mass production. The OC170 was developed first in 1959 for operation at 100MHz, and later the AF114 and u.h.f. transistors like the AF186 with $f_T =$ 600MHz, so that from 1961 to 1967, 30 million alloy diffused transistors were sold.

Germanium was also used for power transistors, the V30/10P for example, capable of 3W dissipation, produced by Newmarket in 1956 and the Mullard OC28 in 1963, the collector current of which was 15mA. The essence of the art of making power transistors was to keep the thermal resistance between the active region of the junction and the case as low as possible so that heat could be dissipated easily by a heat sink on which the transistor was bolted. However, it was clear that for most applications silicon would be the best material. It is perhaps ironic that at this time large contracts were being given to manufacturers in the States by the U.S. Government to set up substantial production facilities to support projects such as Minuteman and other defence programmes, whilst at precisely the same time the U.K. Government was abandoning the idea of a U.K. based nuclear deterrent so that similar British projects were not forthcoming and manufacturers in this country were not so actively encouraged to establish manufacturing plants. These American plants were large, because at that time the yield of good transistors from semiconductor chips was small, calling for a number of parallel production lines. As yields became greater, the manufacturing potential of the plants rose. Thus the U.S. production scene prospered whilst development at this critical time in Britain was much slower.

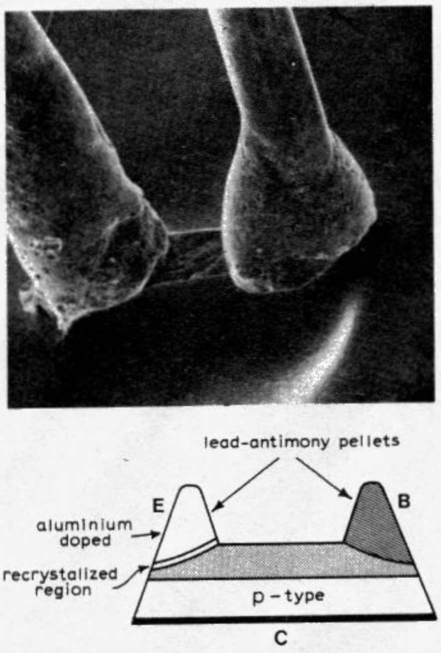
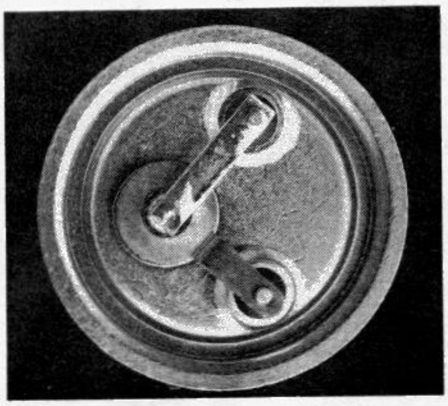


Fig. 2. Cross-section of a p-n-p alloy diffused transistor. Two 100µm wires are soldered to two lead-antimony pellets. The left-hand pellet also contains a small quantity of aluminium, applied as a paint after an initial alloying cycle. After subsequent heating to complete the alloy the left-hand pellet forms the emitter. The other lead is for the base. (Photo: Mullard Ltd)



Header of an OC28 power transistor. The semiconductor chip is towards the left of the header. The longer strap connects to the emitter. The base strap (at the bottom) carries the chip which is about 4.5mm square.

Of course all this resulted, in time, in a substantial cut-back in prices. The *Financial Times* of 27th March 1958 stated that a typical price for a transistor in 1956 was £3, £1.75 in 1957 and £1.4 in 1958 (expressing the figures in new currency). A letter of about the same time from Pye to the Radar Research Station, then at Tolworth Rise, Surbiton, gave the price of an audio transistor, for large quantities, as 80p. All this was but a foretaste of things to come ten years later.

Risks of the game

The end of the 1950s left manufacturers still looking for higher frequency and power, but some of them were by now particularly conscious that the major outlet for transistors would be in data processing. Hence these companies concentrated on faster switching transistors and, incidentally, changed the whole outlook of the electronics industry from being dependent on the fortunes of the communications industry, as had been the case prior to 1939, to being dependent on the ups and downs of the computer industry as is predominantly the case today. Patents covering transistors had been filed on behalf of the Bell Telephone Labs. and any structure which looked as though it would not be an infringement of these patents was particularly attractive, since there was such a large market potential. A number of these cases have been before the courts since.

No discussion of switching transistors can omit reference to gold doping. The use of gold as a dopant had been known from experience with diodes. The presence of gold reduces the lifetime of minority carriers in the collector region and thus reduces the turn off time of the transistor. However, its presence can reduce lifetime in any region of the transistor, including the base region where it is not wanted. The process which is used for most switching transistors is one of diffusion followed by rapid quenching. The diffusion parameters are somewhat critical, hence the yield of devices tends to be reduced by gold doping.

Research being carried out by W. E. Bradley of the Philco Corporation under a U.S. Navy contract had resulted in a fundamentally new type of transistor—the surface barrier transistor. It depended on the properties of the surface of a uniform germanium crystal being different from that of the bulk material. The production method consisted of etching a germanium slice from both sides with a metal salt solution through which current was passing. Then by reversing the current flow, electrodes could be plated on to the germanium. These electrodes not only made contact with the n-type germanium but provided a suitably high density of holes for the device to operate. Bradley's original paper, in late 1953, mentions a frequency of 60MHz and, if this was not enough, it was whispered that this owed nothing to Bell Labs patents. Thus the surface barrier transistor seemed at that time to be a highly saleable commodity.

Philco was a company of some repute and the second-largest U.S. radio manufacturer pre-1939. Their interest in semiconductors had extended to taking part in the Bell Symposium in 1952 which was the first opportunity companies had to "buy-in" on the results of Bell's research. Records for 1955 show that Philco was one of the top three U.S. transistor manufacturers with 70% of the American h.f. transistor market. But Philco were looking for a partner and the company with whom they linked was Plessey. Thus in 1959 the jointly-owned company, Semiconductors Ltd, was set up at Swindon. In addition to the new transistor, Philco brought to the partnership an automated production line and the know-how to run it—and this at a time when other companies were still talking about "green fingers". Plessey were soon disenchanted with the process and found that it was only automated when graduate-controlled—an expensive operation, However, they bought out the Philco interest and adapted the electrochemical process to plate, not just electrodes, but p-type collector and emitter regions to the etched base; the transistor was sold as the M.A.D.T.—Micro-Alloy Diffused Transistor. By 1967, Plessey's interests were growing in other processes using silicon. They decided to cease manufacture of discrete transistors, the company was closed and the whole process abandoned. Philco stayed solely in the germanium market and made no efforts to develop a silicon process. Each year sales and profits fell, until the company was taken over by Ford in 1961 as Philco-Ford. It was finally closed in 1969, much of its production and test equipment being sold to General Instrument Microelectronics. The disappearance or virtual disappearance of companies like Philco, who were leaders just after World War II, shows the heavy cost of bad management decisions or technological mistakes, often leading to an inability to attract and keep good researchers and other key staff.

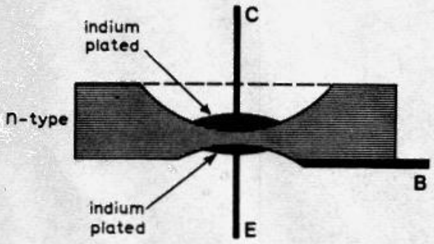


Fig. 3. Microalloy diffused transistor etched by liquid directed at both sides of the slab. By reversing the polarity of the etching current a suitable impurity could be plated, so that the transistor was produced with precise control of physical dimensions, such as the base width.

Silicon takes over

If 1953 was the "Year of the Transistor" as the American magazine *Fortune* proclaimed in an article recently, 1960 was the year of silicon. The Post Office had carried out a study on the accelerated ageing of germanium transistors, and, as a result of this, it was definitely decided that future C.V.D. contracts should concentrate on the use of silicon. S.T.C., Mullard and Ferranti were making silicon transistors. Research was going on at the Services Electronic Research Laboratory at Baldock to make silicon mesa transistors.

In this process, an n-type silicon slice had a p-type layer diffused on to one face. Part of the face was then protected with a photoresist and an n-type layer diffused into the p-type region to give an n-p-n transistor. Finally, the active region of the slice was covered with resist and the uncovered parts of the diffused layers etched away, so that when the resist was removed the transistor was raised up above the remainder of the slice. Hence the name, mesa, after the shape of the hills around Mesa in Arizona, U.S.A., which this profile somewhat resembles.

The process was attractive since it was entirely carried out on one side of a silicon slice. It was soon seen, however, that this was no more than a further step on the road to success. The final etching to make the mesa which controlled the dimensions of the transistor was eliminated leaving the device with an entirely flat surface—the planar transistor.

Ferranti were making the ZT20 in 1960, the first European-made silicon transistors, and S.T.C. following in 1961. The ZT20 was made on 1in silicon slices, later diced into 0.4mm square chips of which 0.13mm was the length of the active area. Transistors like this were made in batches of about 2000 on a slice. A process well suited to mass production was now available.

Epitaxy

The fact that the diffusion of planar transistors was entirely carried out on one face of the silicon slice was at the same time an important advantage and a drawback of the process. Whilst it made mass production a reality, it also meant that collector material of high resistivity had to be used so that there was the resistance of an appreciable mass of silicon between the collector contact and the active collector region near the base. This was a drawback for operation at high power and also resulted in a poorer high frequency performance than had been hoped. Thus even in 1962 S.T.C. could continue to sell germanium tunnel diodes and similar devices as high speed logic elements capable of 50MHz operation, despite all their inherent disadvantages. The solution to this problem was the use of epitaxy.

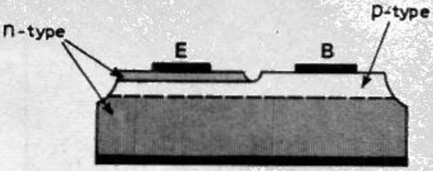
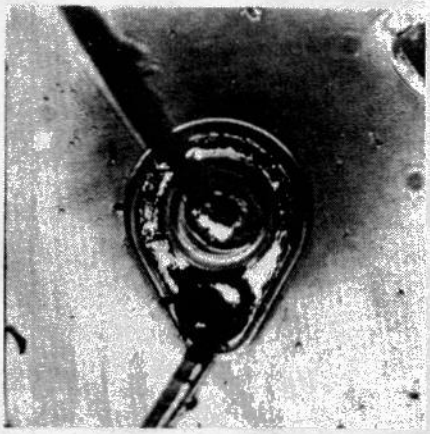


Fig. 4. Mesa transistor, produced by selective masking, diffusion and etching, carried out entirely on one side of the semiconductor slab.



Silicon mesa transistor designed for high speed switching applications, with a current rating of 200mA and a maximum dissipation of 1W. The chip is 0.4mm square.

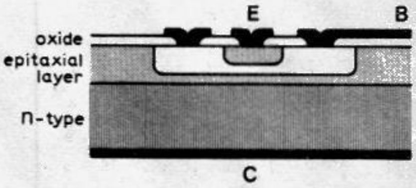
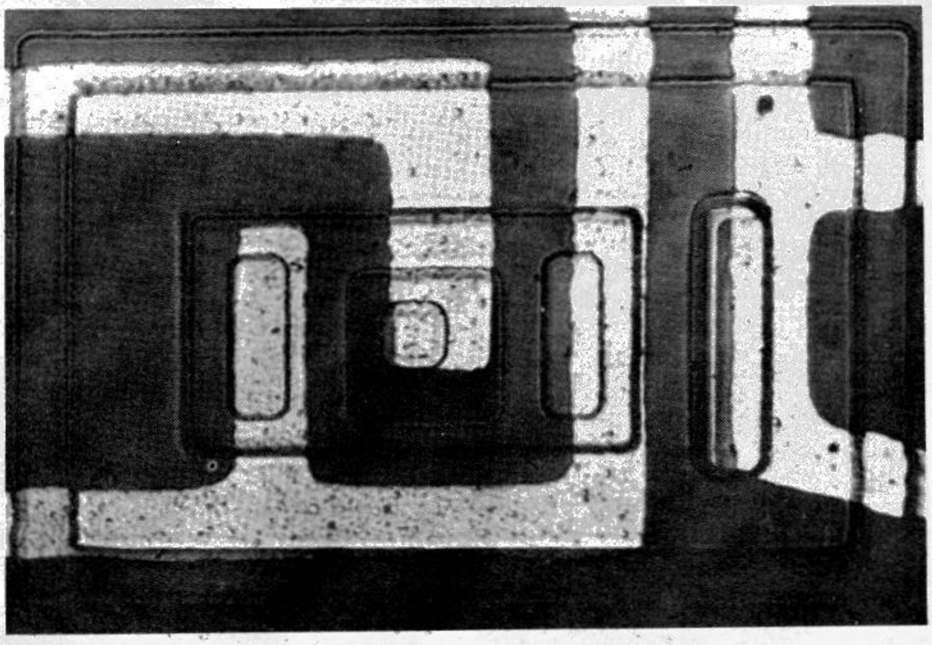


Fig. 5. Planar transistor, like the mesa, produced in one side of a slab of silicon, but with greater control of parameters. The process of epitaxy although first applied to mesa transistors was more fully developed with planar devices.

In the epitaxial process a layer of high resistivity silicon, perhaps 1Ω, was first of all laid down on a much lower resistivity substrate material, perhaps 0.001Ω.cm. The transistor was then diffused with selective masking by photo-resist into this epitaxial layer. Such devices are sometimes referred to as triple-diffused. Although the epitaxial layer had to be sufficiently thick to contain the successive diffusions of the transistor, clearly the bulk of the substrate material is now of much lower resistivity. Faster switching transistors of this kind first made their appearance in the U.K. in 1962.

Perhaps the impact of these advances can

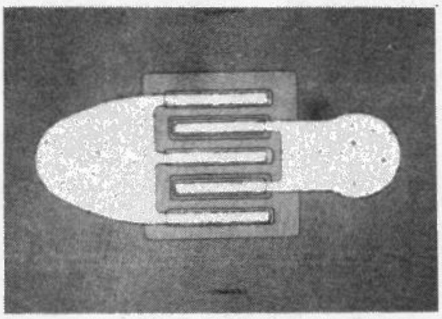


Planar transistor which is part of an integrated circuit (see Part 3 of "The Semiconductor Story"). The picture shows the isolation land and apertures in the oxide mask by which the aluminium pattern can make contact with the collector, emitter and (twice) the base of the transistor.

best be seen by an example. The Mullard BC107, which was a planar epitaxial silicon transistor introduced in 1963, still sells 20 million devices per year. At one time 50 million devices per year were made in the same factory that had formerly made 50 million valves per year. Only now this is done in no more than one tenth of the space formerly used for valve production. This sort of cost-effective approach to the size of production plants was reaping rewards for those U.S. manufacturers who had been supported earlier by defence projects. However, the same market forces which resulted in the evolution of the planar epitaxial transistor had impelled researchers to see that active and passive devices could be diffused into the same silicon slice to form an integrated circuit, but then that's another story, and although it produced severe competition for discrete transistors, it by no means halted transistor development.

Accident and determination

The history of transistor development has been a strange mixture of accident and determination, with high financial prizes at stake. For example, up to 1963 Western Electric, the agent for Bell Telephone Labs had received $8 million in royalties over and above the value of patents and other know-how which Bell must have received from their licencees. Progress was possible because there was early on, particularly at Bell Labs, a good understanding of the physics of semiconductors. It was the gaps in technology, not in physics, which held up progress. The most striking example of this was Shockley's paper in 1952 on the field effect transistor. Shockley and his co-workers had been looking at surface effects when the point contact transistor action was first noticed. Shockley's paper outlines the theory of a unipolar transistor, but it was not until the middle sixties that we could be said to have come full circle, completing Shockley's original investigations, when field effect transistors became a reality. A textbook published in 1964 carries an account of some of the first types which were available commercially. In one form of the transistor a current flows in a semiconductor "channel" between two terminals, called by Shockley the source and the drain, with a reverse-biased junction between this p-type channel and an n-type region which he called the gate. The potential on the gate controls the width of the channel and hence its effective resistivity and thus the current which can pass. Field effect transistors and the later metal oxide transistors in which the gate control was effected by an insulated metal layer instead of the field of a reversed-biased junction, have high input impedance, unlike conventional bipolar transistors, as well as low noise figures, since they have majority carriers and there are no potential barriers. So simple amplifiers were possible, with voltage rather than solely current gains and in some cases capable of high frequency operation.



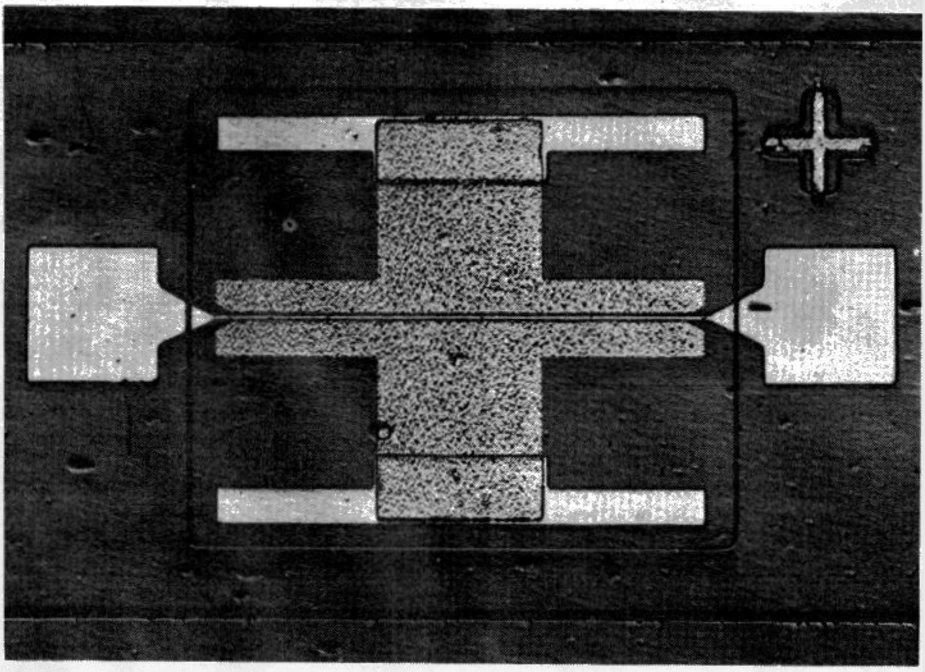
British Post Office silicon planar transistor type 10A2. This is used as the input stage of the submerged repeaters assembled by I.T.T. f_T = 1GHz. The active area is approximately 1mm square. (Photo: British Post Office)

The slow progress with field effect devices was due to technological difficulties such as surface contamination resulting in sudden catastrophic failure, and the high input resistance which enabled electrostatic charge to be picked up in casual handling, which then destroyed the insulation. At length the difficulties were overcome and although field effect transistors have never completely replaced bipolar transistors, for some applications, like high input impedance amplifiers, they have proved particularly successful. One application in which the high reliability of transistors was early seen to be appropriate is that of underwater repeaters for submarine cables, but nevertheless it was not until 1965 that the Post Office and S.T.C. (which became I.T.T. Semiconductors in 1966) could claim that they had products with a sufficient quality assurance for this particular application. The Post Office's 4A series of silicon planar transistors had f_T = 400MHz and the 10 series which followed it had f_T = 1GHz.

The standards of reliability which the Post Office imposed both on itself and on those companies working with them were fantastic. Statistics obtained from accelerated ageing were to give an assurance that there would be no catastrophic failure in service in 25 years, the current gain of a device was not to change by more than 50% and the average change in gain during this period was to be less than 3%. When one realizes that the transistors were to be used in communication amplifiers with high overall feedback so that the gain was fairly constant anyway, this is truely remarkable. Cables like CANTAT in the North Atlantic use these transistors. To obtain this quality assurance, the accelerated ageing tests use 70% of each batch which passes initial screening tests, leaving a mere 30% for submarine service. CANTAT 2 which goes into service in 1974/5 will use 2800 transistors in the amplifying paths alone and work is now going on to extend the 1800 channels of this type of cable to 4000 circuits with transistors having f_T = 3.5 – 4GHz. This is one field where Britain leads the world.

Gallium arsenide

In the search for higher operating frequencies it has long been recognized that gallium arsenide might with advantage replace silicon. Its melting point is 1233°C, slightly lower than that of silicon, but gallium arsenide transistors can operate at temperatures of up to 350°C, considerably higher than silicon. Its mobility for electrons is about three times that of silicon and higher than that of germanium. Mobility is a measure of the velocity with which a charge moves in the material under an applied field and hence high mobility implies suitability for high frequency use. Plessey, for example, started work on gallium arsenide transistors in 1965 and a year later had developed a process for producing an epitaxial layer on bulk material. The discovery of the Gunn effect at this time was



Gallium arsenide n-channel field effect transistor type GAT2 intended for low noise r.f. amplifiers in the X band. Noise figures are 2dB at 1GHz and 5dB at 5GHz. The gate length is 2μm and the separation between source and drain is 6μm. The epitaxial layer is 1μm thick. (Photo: The Plessey Company Ltd.)

another spur to the development of this technology. Once again C.V.D. came to the rescue and gave Plessey a contract to develop a gallium arsenide field effect transistor. This resulted in an amplifier with a gain of 10dB at 12GHz.

No mere replacement for valves

Where have we got to in the search for higher power, higher frequency and higher reliability? Transistors are available today to give 100W at 30MHz in single sideband operation, or 50W at 500MHz or 15W at 1GHz class C operation. These are all silicon transistors and are made in relatively small numbers. The mass production areas are for lower power devices, indeed in some cases they have been the subject of overproduction. The price of the silicon BC108 for example, dropped some months ago to one third of its price in 1969 due to this, although now its price has risen slightly as production levels are adjusted and inflation catches up with the market. Germanium transistors are still being manufactured and one company at least, Newmarket, has declared its intention of acting as residual supplier for the Pye-Mullard-Philips group until at least 1980.

One should not imagine that the transistor merely replaced the valve. Probably the principal effect of the development of the transistor has been to overcome the "tyranny of numbers". It was almost impossible to use valves in large numbers in a system—they consume considerable power for heaters and anode current, they are relatively bulky and inherently unreliable. The transistor has led to the use of modular circuits and sub-systems which often use very large numbers of transistors. Computers, radio communication and radar systems, electronic telephone exchanges of the complexity and reliability we have today would all have been impossible with valves. It is a sobering thought, however, that these developments were to a large extent catalyzed by the U.S. defence and aerospace programme and indirectly by similar competing projects in the U.S.S.R. The fact that 42% of the cost of an intercontinental ballistic missile is accounted for by its electronic equipment is an indication of the weight of the motivation on the providers of semiconductor technology. It is we who must see that, whatever its parentage, the child is a child of peace.

(To be continued with "Solid circuits—a new concept")

Bibliography

1. Shockley, W., "Transistor electronics: Imperfections, unipolar and analog transistors", *Proc. I.E.R.E.*, 40, pp. 1289–1313, November 1952.

2. Shockley, W., "A unipolar 'field effect' transistor", *Proc. I.E.R.E.*, 40, pp. 1365–1376, November 1952.

3. Beale, J. R. A., "Alloy diffusion—a process for making diffused base junction transistors", *Proc. Phys. Soc.*, 70, p. 1087, 1957.

4. Bradley, W. E., "The surface barrier transistor", *Proc. I.R.E.*, 41, p. 1702, 1953.

*5. Kendall, E. J. M., "Transistors", *Pergamon Press*, 1969.

*6. Meyrick, N. L., and Roman, G., "Ten years of semiconducting materials and transistors", *Pye*, 1954.

*7. Scott, T. R., "Transistors and other crystal valves", *MacDonald and Evans*, 1955.

---

*These have excellent bibliographies.

The Semiconductor Story

3: Solid circuits — a new concept

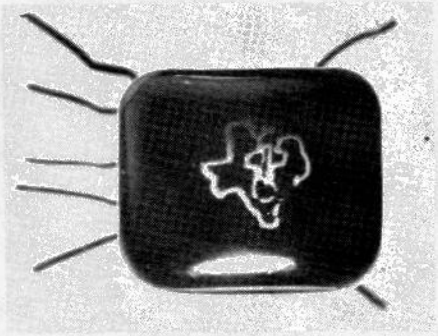
by K. J. Dean, M.Sc., Ph.D. and G. White†, M.Phil., B.Sc.*

The development of the transistor, described last month in part 2 of this series, had been a strange mixture of chance and directed scientific research, of skill with difficult processes and of commercial brinksmanship in which some went too near the abyss and never recovered or withdrew from competing. However, there were occasions when someone intimately involved in the struggle was able to look beyond the immediate technical difficulties and point to an idea not then matched by technological skill, but for which the technology would one day be available. Remarkably enough there are two instances of this happening in the same year, 1952, only four years after the discovery of the transistor effect by Bardeen and Brattain. In both cases the prophesies, for that is what they were, came true in the years to come. W. Shockley, writing in the *Proceedings* of the American Institute of Radio Engineers (now the I.E.E.E.) laid down the theory of the field effect transistor, fourteen years before it was to become a commercial proposition. G. W. A. Dummer of the Royal Radar Establishment (now at Malvern) speaking at a transistor conference in Washington pointed out that semiconductors could be used to make resistors, capacitors, diodes and transistors so that the possibility of putting a number of all these elements on a single piece of semiconductor existed — in fact that it was possible to make an integrated circuit. It was however to be seven years or so before this idea reached any sort of fruition and about sixteen years before these two, the integrated circuit and the field effect device, came together as a complex commercial product.

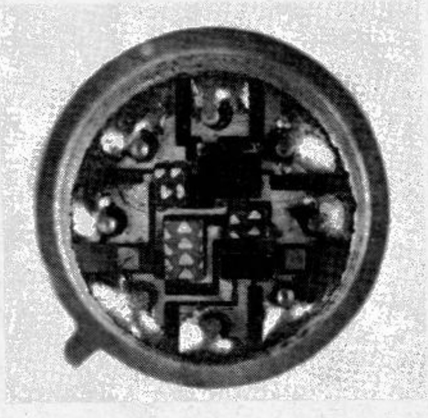
Of course the germanium technology of 1952 was quite inadequate to put Dummer's idea into practice and it was five years before the Plessey Company, who were by then more interested in precise photo-chemical processes, were given a contract in association with the R.R.E. to investigate the possibility of a solid circuit. In 1957 an international symposium on electronic components was held in Malvern at which, reported *Wireless World* in November, the solid circuit was little more than an idea to be discussed in the same breath as ferrite blocks and resin-potted circuits. But there was one point which was significant — the solid circuits being proposed in 1957 were silicon, not germanium.

Technology available

By this time a number of other companies both in the U.S.A. and in Europe were interested in solid circuits, amongst them Texas Instruments (in Bedford as well as in the U.S.A.) and Fairchild. Not only were silicon transistors available but the mesa process had also been recently developed, largely by Texas Instruments. Now this process has the important advantage of requiring diffusion from only one face of the silicon slice. Hence it was thought possible to place various active and passive components side by side on a single slice and then inter-connect them. In 1958 this is what Texas were able to show they could do. As was the case with transistors where increasing skill with technology and governmental patronage produced a variety of transistor types, changes in solid state techniques had a vital impact on the development of integrated circuits. The key technology was the development by Fairchild of the planar process, so that even by 1960 it was clear that planar devices would most easily lend themselves to interconnection as solid circuits. In fact it can be argued that two of the major efforts since that time have been to minimize the profile contours of silicon chips and reduce the size of transistors within the chip. These have been brought about using modifications of the planar process.



Early Texas mesa integrated circuit showing the vitreous enamel package, from the underside of which the connecting leads protrude.



Multi-chip integrated circuit by S.T.C. mounted in a TO5 header on a printed circuit board on which the interconnection pattern has been etched. The circuit, sold in 1964, is that of a d.t.l. gate.

The patronage which proved decisive and turned, alas once more, a British idea into a foreign product, came from the U.S. Government. The Minuteman project was at the end of the '50s the American contribution to the U.S./U.S.S.R. arms race and represented the ultimate then possible in electronic sophistication. It was funds from this project, principally to Fairchild but also to Texas which provided the immediate incentive to devise high component-density circuits of great reliability for use in the limited space and very difficult environment of a missile. Thus the early integrated circuits were born. Although by this time a technology to make a form of integrated circuit was available on a laboratory basis, it had a number of limitations, both of cost and as a production method. Failure to produce a reliable isolation technique meant that multi-chip circuits were the best that many companies could do. One chip might carry a single transistor, another might have a resistive network and a third might consist of diodes. The chips were at first mounted on a suitable sub-divided printed

---

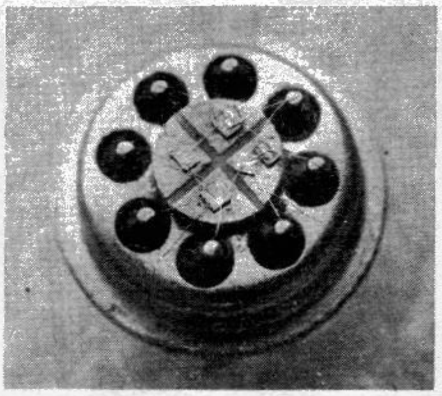
* South East London Technical College
†Twickenham College of Technology

circuit board or a ceramic button which was in turn mounted on a TO5 header. Another area of difficulty lay in the interconnections. Contact pads were provided on the silicon chips by depositing an aluminium pattern. This metal had low resistance and was found to give good adhesion to the surface of both p-type and n-type silicon. The interconnection leads were gold wire. A reaction may take place between these two metals at the fairly high temperatures used in bonding. This results in high resistance purple or black coloured compounds known as purple plague and black death respectively. These are intermetallic compounds which arise in the presence of silicon. Whilst purple plague for example could and did exist with discrete devices, it was with early integrated circuits, particularly multi-chip circuits, that it became widely known. For instance some of the i.cs used in Minuteman missiles had a gold/aluminium interconnection system and so suffered from purple plague.

The development of integrated circuit technology was almost rocket-like with the U.S. Government support it attracted. The 1962 Minuteman II project might be regarded as the second stage of the rocket. In this project Texas had a contract to supply 300,000 i.cs thus setting the scene for later large scale production. At this time the U.K. Government was abandoning its independent nuclear deterrent; for example, the Black Knight programme was cancelled. There was therefore relatively little incentive in Britain to develop British i.cs for the world-wide defence market which undoubtedly existed. The state of the art here can be seen in a report from S.T.C., "Report on commercial valve developments — solid state circuit techniques" which showed separate resistors, capacitors, transistors and diodes on a single chip wire bonded to give an r.c.t.l. gate and stated that the first circuits were made in March 1962. Similar gates had been available in limited quantities in the U.S.A. for at least 12 months before that and in 1962 the first commercial planar circuits were already being advertised in British journals by Fairchild. These were r.t.l. (resistor-transistor logic) gates and were capable of operation at 1MHz. The chips were typically 1mm square — one hundredth of the surface area of the ⅜in square chips proposed by Dummer ten years earlier.

Why logic gates as i.cs?

The first integrated circuits were almost exclusively logic circuits. This was because the electronic control of missiles was very largely of a digital nature and because it was much easier to design switching circuits which had only two states of operation than it was to produce linear amplifiers. Silicon technology made it possible to design circuits in which the tolerance was relatively tight between components in a circuit but it was unable at that time to yield circuits in which the absolute tolerance of any one component could be kept small. This suited the design of switching circuits.



Multi-chip low noise cascode amplifier for use at frequencies up to 100MHz produced by Marconi Microelectronics in 1965. The chips are mounted on a ceramic button fixed to the header, and are wire bonded; the button has been divided into four "lands".

One of the transistors from a low noise 100MHz cascode amplifier. The transistor was made by an early form of planar process and the chip size is 0.46mm square.

Perhaps stage 3 of the "i.c. rocket" was fired when it was realized that the limited market of defence requirements could be replaced by the much larger market of the growing computer industry which also used logic circuits and already had very definite views about their modular nature. Thus from the start the need for switching circuits rather than linear amplifiers was paramount. Integrated circuits were therefore gold doped and this method of obtaining a speed advantage was almost always followed until the advent of Shottky diodes in about 1970. By the end of 1963 work in the U.K. was catching up and a report on a C.V.D. project by C. P. Sandbank describes the manufacture of circuits which include isolation lands, just as are used in the epitaxial circuits we have to-day, with buried layers to eliminate parasitic p-n-p transistors, and of course with gold doping. These circuits were a form of transistor-capacitor logic with 35ns propagation delay through each gate. This at least was an improvement.

Industrial pressures

The effect of all this was a scramble for a place in the market, and a highly competitive market it turned out to be. Fairchild linked up in Europe with the Italian company, S.G.S., later to separate again. Elliott set up a production line at Boreham Wood and Marconi at Witham near Chelmsford and at Glenrothes in Scotland. These companies later merged with G.E.C. who had been in semiconductors from the start, and with A.E.I. who had already withdrawn from making small junction transistors. Eventually the manufacturing plants at Glenrothes and Witham were closed although not until near the end of the sixties. Meanwhile Plessey's semiconductor plant was turned over almost exclusively to i.c. manufacture at the expense of transistors. What brought about these traumatic changes?

In 1960 integrated circuits cost £20 per package and were available for military purposes only. Ten years later they had fallen to one per cent of their original cost and were incorporated in a wide range of industrial equipment and were even making a substantial impression on the traditionally cost-conscious domestic market. To understand how this came about one must know something of the factors which influenced industrial growth and falling price. Circuit development costs are substantial when only a small number of devices is required. It was commonly stated in the middle sixties that the price of the design work for a set of masks to diffuse an integrated circuit was £10,000, but frequently all this cost had been covered by defence contracts and it did not recur so long as the same device was manufactured for the industrial market. Labour costs are high when production lines have to be staffed with costly graduates, but as the technology becomes better understood less skilled labour is employed, so that eventually plants were set up in low labour cost areas such as Taiwan, Hong Kong, New Guinea and Portugal — often referred to as "off-shore" plants. Due to the small size of i.cs air transport charges are very small and slices could be diffused centrally under excellent supervision and good environmental conditions and then flown to an off-shore plant for encapsulation and testing.

The cost of the material used in making an i.c. is directly related to the yield of good devices which can be obtained.

The resistivity of slice material affects the tolerance of components from one chip to another. It is now possible to hold this to less than 15% instead of 25% formerly from the centre to the edge of the slice. The number of dislocations in the material was typically 30,000 per sq. cm. It is now only 500 per sq. cm.

Circuit designers soon realized that active elements took up less space than the passive elements which they could replace. The space occupied by a 1kΩ resistor was at least equal to that taken up by four transistors and in some cases more. If a large value resistor could be eliminated or its resistance drastically reduced by using a few transistors this could well result in a smaller surface area for the circuit. Hence a new circuit design philosophy developed in which active elements were to be preferred to passive ones and resistors were restricted to values between about 30 Ω and 1k Ω.

The first commercial integrated circuits which were monolithic, i.e. on a single chip, were produced using one inch silicon slices. To-day 2in slices are typical. So not only did percentage yields rise, but device sizes became less and four times the number of devices were provided on each processed slice. The high percentage yields and the greater throughput per slice meant that the manufacturing plants which had been set up with U.S. Government money were in an extremely strong position to compete not only with the U.S. manufacturers but with manufacturers in Europe, in Japan and even in the U.S.S.R. In the Soviet Union the effect of the large American output was to concentrate effort on thick film and hybrid circuits, for silicon circuits could always be imported, for example, through Austria and Hungary. Thus, by 1970 there was only the barest token U.S.S.R. export market for integrated circuits and then only in specialized circuits with only a very small market potential.

High speed circuits

As the computer industry became more and more the main customer for integrated circuits, so "he who pays the piper" began to call the tune. Computer manufacturers wanted two things: reliable high speed circuits and the availability of the same package from a number of sources. The i.c. suppliers felt bound to comply and so with "second sourcing" available a new twist was given to competition and price cutting.

In late 1963 the American Motorola company started to make emitter-coupled logic gates. A typical dual two-input NOR gate in an eight-lead TO5 can consisted of a single 1mm square chip. These Motorola e.c.l. gates had propagation times which were less than 5ns, but they were potentially even faster and by 1971 types called MECL3 with less than 1ns delays were amongst the fastest circuits on the market. One of the problems of the Motorola e.c.l. which delayed acceptance of these gates was that the switching potentials were less than 1V apart and neither of them was at the potential of either supply line. This necessitated the use of special reference voltage i.cs in addition to the gates of which a system might be composed.

In 1965 both S.T.C. and Marconi made agreements to second source Fairchild diode-transistor logic and Plessey started to make i.cs at Swindon on a production line basis — a year later they were producing 300,000 circuits per annum. Though Plessey had started in r.t.l. they were now second sourcing the Motorola MECL series. Well separated logic levels and under 20ns propagation time were provided by d.t.l. gates. Their competition in 1965 was from r.t.l., the natural successor of modular circuits such as Norbit and Minilog. They had been developed early and their design costs amortized, hence they were cheaper then than other i.cs. But r.t.l. gates were not fast enough for the computer manufacturers, so production quantities fell and prices no longer fell. Principally developed by Texas, t.t.l. soon took over with under 10ns propagation time and in due course were second sourced also, by Mullard, Siemens, and (in 1968) by I.T.T. (formerly S.T.C.).

Dual-in-line packages

The first Texas mesa solid circuits were encapsulated in vitreous enamel and subsequent circuits used standard transistor packages, principally TO5, whilst military buyers insisted on more expensive flat-packs. However, these became less acceptable as more complex circuits were devised which required more connecting leads. Hence dual-in-line plastic and other hermetic packages were progressively introduced from 1966 until eventually this type of package became exclusively accepted for industrial and commercial applications and for many military purposes also. In fact, once adopted, it has been used for a number of other non-semiconductor electronic components.

Linear circuits

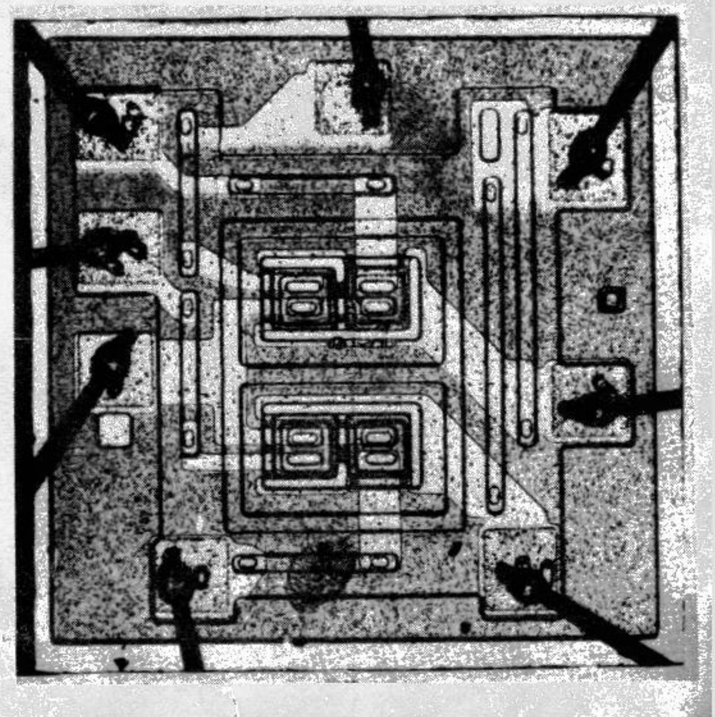
As familiarity with the technology grew it was certain that at some stage linear circuits would be tackled by leading manufacturers. In the U.S.A. Fairchild introduced the 700 series of operational amplifiers and in the U.K. Plessey was working along similar lines. One of their SL500 series, for example, designed in 1967, had a current gain of 26dB at 30MHz with a response from 5MHz to 100MHz. The amplifier circuit elements consisted of three transistors, a diode, seven resistors and three capacitors on a single chip 1.5mm square. The advantages of excluding capacitors, which were wasteful of surface area, were quickly realized. Gain and frequency response were then controlled by external reactive shunts and by internal feedback. Thus one amplifier circuit could be used for a wide range of applications with consequent sales benefit to the manufacturer. For example, the Fairchild μA730 was a differential amplifier which consisted of seven resistors and five transistors whilst the μA709 high gain amplifier which used external feedback to control gain consisted of 15 resistors and 15 transistors. Neither of these circuits included areas specifically devoted to capacitance. Of course, capacitance is always present, limiting frequency response.

By 1968 a number of manufacturers were experimenting with special purpose i.cs hoping to break into new markets, such as radio, television, automobile electronics and the "white goods" trade (fridges, washing machines, etc.). One company produced a car radio chip — basically a superhet with reflex i.f. and a.f. amplifiers though they found a market for it very difficult to obtain, and in 1969 Plessey had a single chip colour demodulator for Rank-Bush-Murphy colour television receivers.

A goal is reached

It was in 1968 when those two great ideas, mentioned earlier, of the solid circuit and the field effect transistor came together, and none too happy a union it was at first. Field effect transistors had been in production since about 1963, first with junction gates and later with metal oxide insulating gates. However their reputation for reliability was very poor.

Small silicon area, fewer diffusions than for bipolar transistors, high input resistance and a high fan-in when used as switches are all properties of m.o.s. devices. They depend however on surface effects and so are liable to surface con-

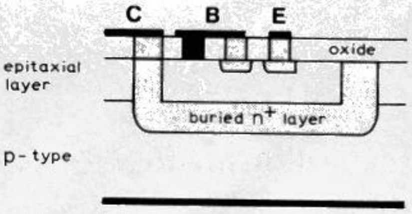


Motorola MC910. A 1mm square chip containing four transistors and six resistors. Designed in late 1963, this is an r.t.l. dual two-input NOR gate which has a 40ns turn-off time. The four input resistors are 1.5kΩ each.

tamination, and in some cases suffer from poor surface stability. Catastrophic failures were not at all uncommon due not only to contamination but to electrostatic pick-up. Also silicon dioxide decomposes in the presence of aluminium resulting in pin-holes in the insulation layer which are fatal to the device. Contamination is particularly likely when an m.o.s. circuit is encapsulated in plastic, such as the dual-in-line packages then becoming popular. The solution to this problem was found to be to use not only an oxide layer but also a nitride layer to give passivation of the surface since silicon nitride is not affected in the same way by aluminium as is silicon dioxide. These m.n.o.s. gates were introduced by a number of companies, among them Ferranti. A few months later Plessey set up a production line for m.o.s. circuits at Swindon and by 1972 30% of their output of 1.2 million chips per annum consisted of m.o.s. circuits.

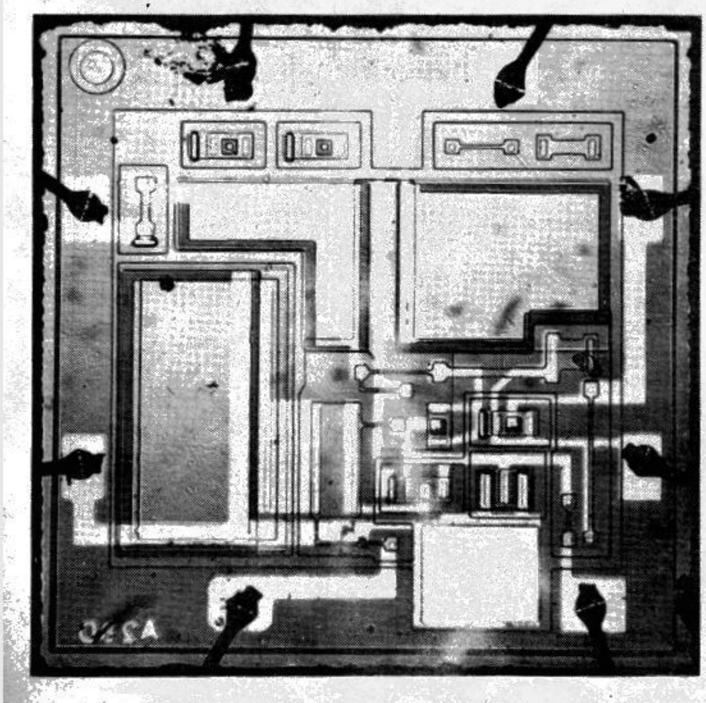
Finding their feet

Soaring yields, even with integrated circuits, due to familiarity with processing technology, and the lure of even larger bipolar and m.o.s. circuits all enticed manufacturers to do better and build bigger, while growing competition and the dramatic failure of some and ever falling prices were never far from their thoughts. Of course chips have got bigger. A typical maximum chip size in production now is 4mm square with the occasional 6mm square "special" but m.o.s. has been something of a disappointment with the larger chip sizes. There has been steady progress towards m.s.i. (medium scale integration) and l.s.i. (large scale integration) except for agreement on exactly where a function on a chip becomes large enough to warrant the term m.s.i. or even l.s.i. (more about this in part 4). However, there have been some interesting ideas floated by engineers about methods for making larger circuits without yields becoming vanishingly small and it may be that among them are those who have had a glimpse of what the future really holds.



Collector diffusion isolation (c.d.i.) process involves diffusion of an n^+ layer into a p-type silicon slice. Subsequent growth of a p-type epitaxial over the now buried n^+ layer is used to hold emitter diffusion and collector diffusions which link up with the buried layer. This results in the isolation of each transistor so formed.

As long ago as 1966 the theory was being proposed that circuit yield depended on the density of the interconnection pattern of the aluminium on the surface of the chip. It was claimed that devices on the chip could be made smaller and the separation between devices less so that the limiting factor in the technology was the resolution of the aluminium pattern. Hence ways were sought to reduce the number of conductors on the chip. Some diffused layers were conveniently available as underpasses, but generally an underpass takes up more space than the corresponding conductor on the surface. It was suggested that a number of interconnection layers would reduce the density of conductors in any one layer so much that yields would rise. But more layers mean more masks and yield is proportional to the power of the number of masks, so yields fell when this was attempted. Some, like Fairchild, had a special slant on this problem: the chip consisted of say, 32 gates, and the first layer of metallization connected the circuit elements on the chip into gates. The customer was then asked to design the pattern which interconnected the gates to form the functions he wanted. The idea foundered, both due to the low yields which meant that prices were high, and because customers did not see why they should do part of the i.c. manufacturer's work for him. It is interesting that those manufacturers who either do not use two-layer metallization or who have tried and failed point to the contour of the silicon surface as the core of the problem. Although one might imagine the planar surface to be flat its profile is far from this with windows in the oxide layer making contact with the various diffusions and aluminium contacts as well.



British-designed linear amplifier chip which shows the area of surface taken up by capacitors, compared with the much smaller areas taken up by transistors and resistors. The circuit is that of a capacitor-coupled r.f. amplifier.

Beam leads and flip chips

High on the list of advantages of solid state devices is reliability. It has long been recognized that the weakest link in transistor technology is the bonding of the chips to the posts on the header, or directly to a printed circuit board or other components. Two alternatives to wire bonding are available. Flip chips have thickened bonding pads so that the device can be bonded face to face by contact between these pads and another surface. Although some work has been done with flip chip i.cs, automatic assembly of chips is seldom used so that it is chiefly with discrete transistors that they have been used. The main exception is the I.B.M. solid logic modules where flip chips have been used in assembling random access memories.

Beam leads are produced by multiple deposition, usually of platinum and gold to extend the conductors beyond the edge of the chip, so that when the silicon is etched away from the edges of the chip, the beam leads protude. Beam leads have been found more appropriate for i.cs, the circuits being separated either by a lap and etch technique or by air abrasion. Beam lead technology is of some significance in the U.S. but once again there have been no significant contracts in the U.K.

What next?

An early Texas project had been called "A computer on a chip". At the time it seemed this was just American "talk" but l.s.i. has now turned this into a potential reality. Part 4 of this series will look at the development of l.s.i. and make some sober guesses about the future. Some cynics might wonder whether there is a future, for the last five years have certainly brought over-production and shown the perils of being tied to a pace-making industry like computing. But the faint-hearted don't work in semiconductors. If the cost of a small domestic car like the Vauxhall Viva had fallen as dramatically as that of integrated circuits over the last ten years, its cost today would be comparable to that of a secondhand bicycle.

to be concluded

The Semiconductor Story

4: Large scale intentions. Conclusion of a series of articles commemorating the 25th anniversary of the transistor

by *K. J. Dean**, *M.Sc., Ph.D.*, and *G. White†*, *M.Phil., B.Sc.*

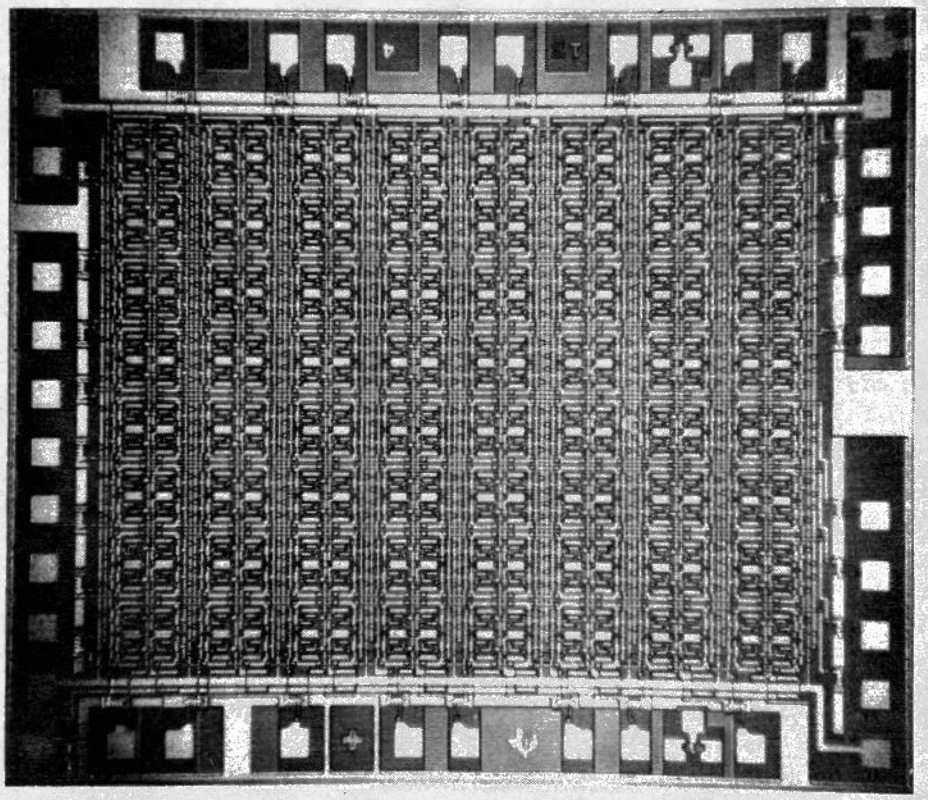
Since 1945 the industrial society in which we live has been one where technological change has been the normal state of affairs. It is not easy to plan such changes; indeed there has been very little worthwhile market and technological forecasting. Our national research establishments have been involved in bringing changes about, but it does not seem to have been a part of their role or that of industry to formulate clear research and development goals based on market assessment. To a surprising extent the semiconductor industry has been a victim of circumstances rather than their master. Its fortunes were founded on the arms race and further encouraged by the U.S. space programme. Again we have seen that military confrontation seems necessary to bring about major scientific developments. (There must surely be some other way.) Defence contracts helped establish large production plants when yield efficiencies were small, so that increasing skill and consequent falling production costs brought overproduction and "dumping". Fierce competition resulted in casualties despite the larger market which became available. A situation was arising which, though so clearly visible in retrospect, no one appeared to notice then.

Larger chips

As the move to put more electronics on a single chip got under way even greater attention was paid to the problems of increasing yield. There are, perhaps, three golden rules if high yields are to be achieved but, like all such rules, they are easier to state than to implement.

First, the processing should be simple The main difficulty here is with gold doping which is a particularly critical process necessary because charge-storage takes place in the lowest concentration area of doping, which is usually in the collector region of the transistor. Gold doping decreases life-time and so reduces charge-storage. However, this effect can also be mitigated by a diode between collector and base, so that the overdrive current goes through this anti-bottoming diode rather than the collector region. Unfortunately, if a silicon junction diode is used it has the same forward characteristics as the silicon junction transistor which it is trying to speed up. This difficulty was overcome by using the Shottky barrier diode formed by aluminium on the silicon, which has a knee voltage of 0.3V instead of 0.5V for a silicon junction diode. The use of Shottky diodes to clamp a transistor was originally developed by Texas in 1964. In some devices the storage is in the base region. In this case, a second emitter is provided for the transistor on the chip and connected to a Shottky diode to remove the charge. Devices where speed is obtained from Shottky diodes are compatible on the same chip with linear circuits whereas gold doped circuits are not. They are also compatible in the same system, but not the same chip, with similar designs for gates, but which use gold doping. An example of this is the Texas 74 series which has been second-sourced by a number of other manufacturers.

Secondly, the number of stages in the process must be kept small. But as the circuits which industry require become more sophisticated, such as gates with good speed-power ratio and high fan-in and fan-out and capability for wired-OR connections, so the number of stages tends to rise. There are for example typically three masks needed for a single transistor, eight for a t.t.l. gate and ten for some linear amplifiers. Only five masks are needed for m.o.s. gates but here speed problems exist, particularly where m.o.s. gates are interfaced to external connection. The yield of a single diffusion is inversely proportional to the area of the chip. That of a transistor or integrated circuit with *n* diffusions is proportional to the yield for a single diffusion raised to the power of *n*. Thus the yield for planar transistors with three diffusions must be extremely high



Dual 64-bit shift register first available commercially in the U.K. in 1967 and typical of the state-of-the-art at that time. (photo: Texas Instruments)

---

*South East London Technical College.

†Twickenham College of Technology.

before worthwhile yields can be expected from five- or eight-mask integrated circuits often 100 or more times the surface area of single devices. If a large enough system can be put on a chip the interfacing problems are less but even then an m.o.s. system is usually slower than a comparable bipolar one, and in some circumstances this is important. Further, such a solution infringes Law 3.

The third law is to keep the chip area small; but this discussion is all about increasing chip size. Scoring heavily here is m.o.s., since the devices are self-isolating. In any case, one should take care to see that isolation diffusion, sometimes known as "lands", is minimized. One process which leads to smaller devices is ion implantation. Here dopant impurities are implanted by ion bombardment rather than by diffusion. The process is compatible with planar technology and gives good control of junction profile but it is more expensive than diffusion since it depends on vacuum technology and the use of high energy accelerators.

Thus we have seen that with these three "laws" there are ways by which they can at least be bent, if not broken. The extent to which they can be bent and there still be a profitable yield is a measure of a company's success with the technical problems. Thus typical chip sizes for m.s.i. (medium scale integration) is 2mm square for bipolar t.t.l. with about 40 gates on a chip of this size. Somewhat larger m.s.i. chips can be made if m.o.s. gates are involved, perhaps 4mm × 3mm with about 500 gates irregularly connected or, say, 1024 bits of random access memory, the latter being, of course, regularly connected.

New planar processes

Now it can be seen that m.o.s. circuits are simpler and hence cheaper to produce; also they represent higher circuit packing densities than bipolar gates but in terms of performance m.o.s. is often at a disadvantage. Therefore, in 1970 manufacturers began to investigate bipolar processes which seemed to offer prospect of being competitive with m.o.s. For example there was the c.d.i. process (collector diffusion isolation) developed first at Bell Labs and then by Ferranti, the Isoplanar process of Fairchild, the Process IV which was suggested at Plessey's research centre at Caswell, and the Dutch Locos process developed by Philips. All of these were compatible with circuits which could operate in excess of 1.5GHz and all of them had the advantage of using less surface area than earlier processes. The c.d.i. system, for example, started with a slice of 10 to 20 Ω.cm p-type silicon into which n^+-layers were diffused. These were later to be the collectors of transistors formed in a 1 Ω.cm p-type epitaxial layer put down on top of them. Then n^+ diffusions were made through the epitaxial layer to make contact with the now buried n^+ layers laid down at first. These not only acted as collector contacts but isolated the area within as in the photograph of the Ferranti c.d.i. chip. In this base area the n^+ emitter diffusion is made, as well as any second emitter for a Shottky diode. After the oxide has been deposited and holes cut in it to gain access to the electrodes, silicon is grown in the holes to the same level as the oxide, thus giving a flat surface.

Other developments

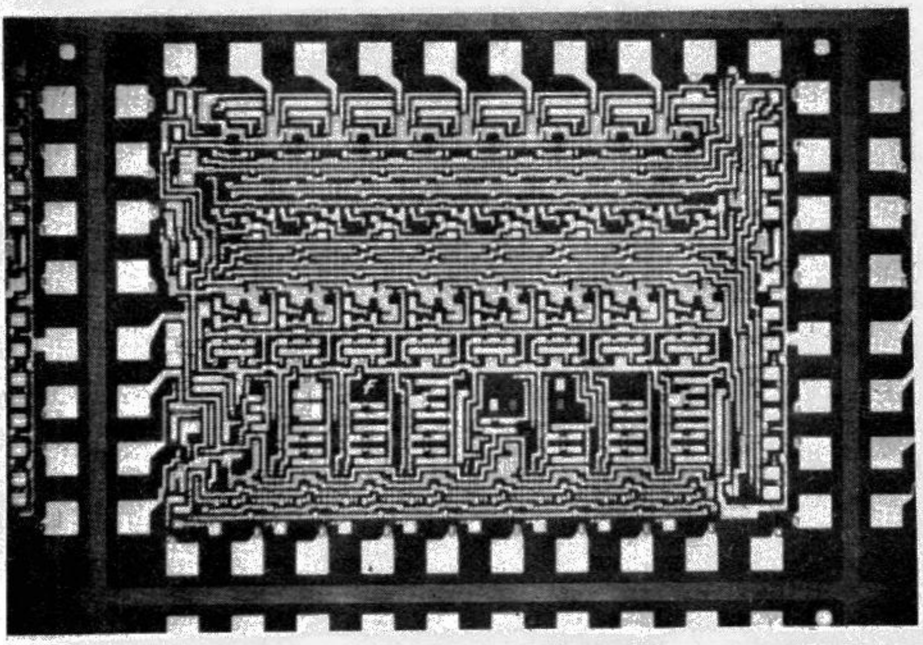
If any semiconductor manufacturer is asked about possible developments he will immediately reply that computer memories, read only and random access types are obvious areas of development. To-day a r.o.m. (read only memory) of about 4 kilobits can be made and this will probably be extended to 32 kilobits in the near future. Random access memories will also be of similar size. Complete processors are being made for the hand calculator market. These are m.o.s. chips since their slow speed is no disadvantage for manual operation. This is a growth area at the moment and prices of hand calculators are falling rapidly. This industry in which the Swiss once had a sizeable share is now dominated by Japan but often with U.S.-designed m.o.s. chips. It is whispered that the Swiss have made careful surveys before joining the competition and have decided that it is now too late to compete. Certainly they have had no indigenous computer industry to help develop their small semiconductor facilities.

What is needed now is to find markets other than in computing. Some possible ones are in communications and in various domestic industries — entertainment, cars and white goods. In telecommunications the first electronic telephone exchange to use integrated circuits was the London empress (01-603) exchange which has 10,000 i.cs. The telephone network will become increasingly digital. It is expected that by 1990 all additions to the network will be digital ones, but 17 years is a long time to wait when you are selling silicon chips. Not only will solid state crossovers be used in electronic exchanges but there is clearly a market potential, maybe for our ailing computer industry, in data processors for telephone exchanges. How often is the engaged tone heard when the number is not engaged; it is the route which is fully committed. Exchange processors are needed to effect re-routing depending on the traffic being carried at that time by a number of exchanges. Visual solid state elements are another area where a start has been made. Plessey had a much publicized chip in 1966 which carried a matrix of 10 × 10 photodiodes. Perhaps we shall see a larger and more closely packed matrix with their Process IV before long, and so bring us a step nearer to replacing vidicons, or at least for document reading.

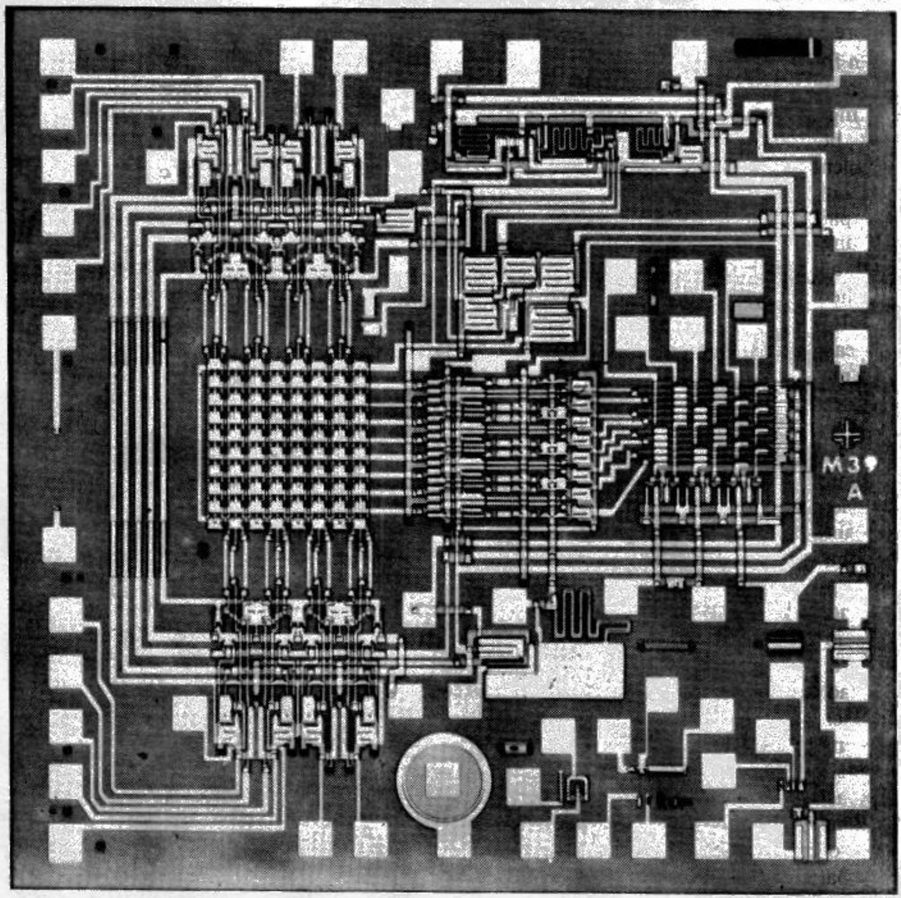
Bell Labs have been working on picture 'phones for some time using m.s.i. chips which supply data to update a store only where picture content is changing, but this is no short-term research project. We are more likely to see low speed facsimile, perhaps augmenting or replacing the national telex network, before this.

British Rail have a large financial commitment in the development of high speed trains. As speeds become higher, say 200 m.p.h., one can no longer rely on the driver to make appropriate decisions in the much smaller time he has available to him, so that here again integrated circuits will find markets in a central processor and its associated control systems.

The white goods market offers prospects for circuits to control washing machines and similar equipment. The automobile market in which four million cars are made in Europe each year has a potential of perhaps £30 per car for electronics, to control ignition and petrol injection and to sample various



Complete serial arithmetic unit on one chip, for use with eight-bit numbers. The chip consists of 200 m.o.s. gates and was designed in the U.S.A. by Fairchild and marketed in the U.K. in 1968. (photo: S.G.S. (United Kingdom))



This chip carries a non-volatile m.n.o.s. 64-bit memory (left-centre) with p-channel m.o.s. decoding circuits which require a 40V signal to drive the memory cells. The storage time for which the memory can be retained in the event of power supply failure and which can be measured in days, months or even years, is a function of the thickness of the oxide and nitride layers. (photo: The Plessey Company)

transducers so as to indicate a fault or warn of any dangerous condition, even of speeding. This represents an enormous market which is virtually untapped to-day. When cars are advertised as "solid state controlled" and the price of hand calculators makes slide rules obsolete, the semiconductor industry will be freed from the tyranny of computing.

What of l.s.i?

In 1968 manufacturers like Texas were proposing to make 2in slices, diffused to comprise the gates of a single system, that is, using a slice as a single integrated circuit. It was then suggested that these slices would be probe-tested and a discretionary wiring applied so as to avoid or short out faulty devices. But individual probe testing on this scale is expensive: in fact the size of chips may ultimately be limited, not by the technological skill with which they can be manufactured but by the time taken to carry out testing, to interpret the tests, and the cost of testing. This is especially true when the chip logic is of a non-repetitive nature.

There is however a problem which is more fundamental, even than this. What is there that is sufficiently complex or so large in terms of circuitry that it warrants the use of a single chip of this size and which at the same time is of such a general nature that it will sell in such quantity, perhaps to a variety of users, to make it an economic proposition? Unless this can be satisfactorily answered l.s.i. cannot be really viable. There are of course some "answers" which might be considered. Random access and associative memories for computers, use so many interconnections between cell locations, for addressing and so forth, that it is desirable to avoid the inter-chip connections which would occur using m.s.i. chips as sub-sections of the memory and joining the sections by printed circuit boards. Also, when processors need to be of small physical size and speed is not an important parameter, as with hand calculators, a single chip has advantages such as minimal wiring and servicing costs. Nevertheless there are but few cases where a convincing argument can be put up, to show that it is not good enough to use beam lead i.cs interconnected as a hybrid system on a thick film substrate. Genuine arguments are in short supply: it is not enough to have large scale intentions.

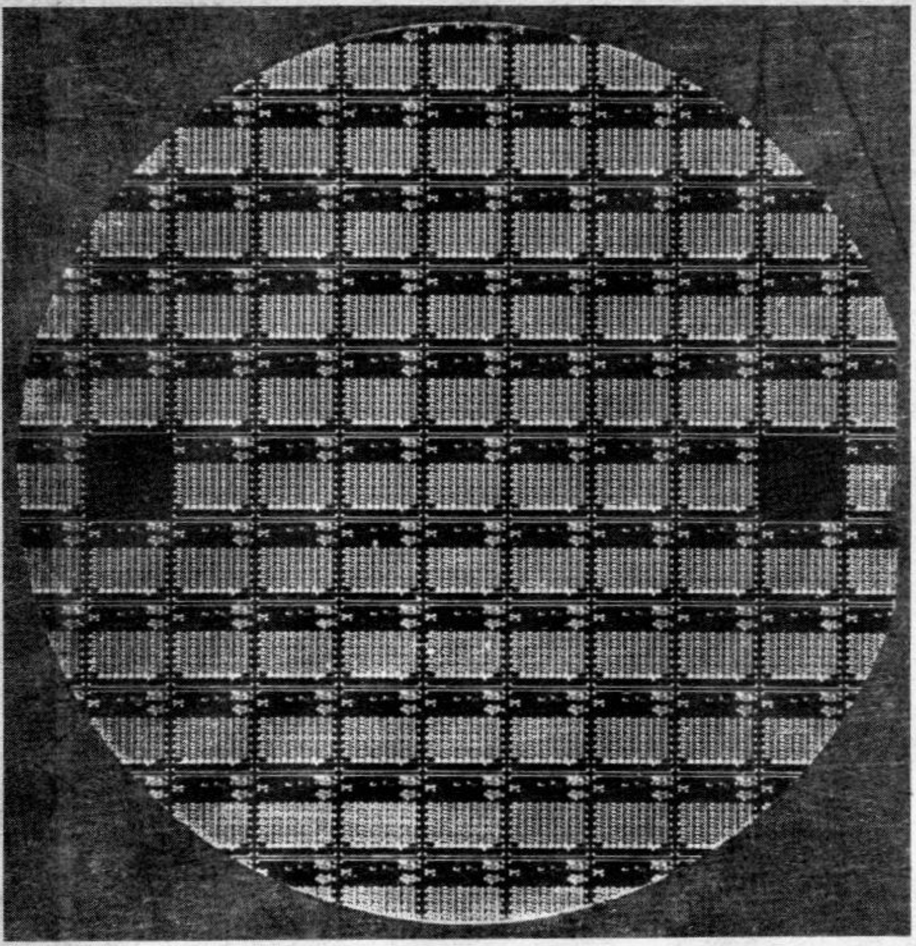
Manpower

Now let us look at some of the significant trends regarding people, prices and prospects. In 1966 the Manpower Research Unit of the Department of Employment (then the Ministry of Labour) carried out a study of the electronics industry, published by H.M.S.O. in 1967, to forecast labour requirements in detail to 1970 and more generally beyond that. It claimed to pinpoint the major growth areas of employment. It forecast for the period 1965–1970 over 50% increase in the jobs available for wiremen and production workers, 41% increase for testers, 35% for scientists and technologists and 32% for technicians. So far as longer term trends were concerned it called for more technicians and for them to have higher qualifications. It did however observe that semiconductor manufacture "would become still more capital-intensive". The predictions were not believed, for negligible support was given by industry to our technical colleges to increase the number of trained technicians by the large figure of one third. There were even cases in 1971 of withdrawal of industrial support for some block release and sandwich students. As for the scientists, it was recently stated that in 1971 only the chemical industry had more unemployed Ph.Ds than electronics. The period 1969-1971 has seen a massive cutback in the computer manufacturing and electronics industries — takeovers, closures, redundancy — and very few in the semiconductor sector making a profit. Perhaps manpower prediction, like weather forecasting, is an area in which we still have much to learn.

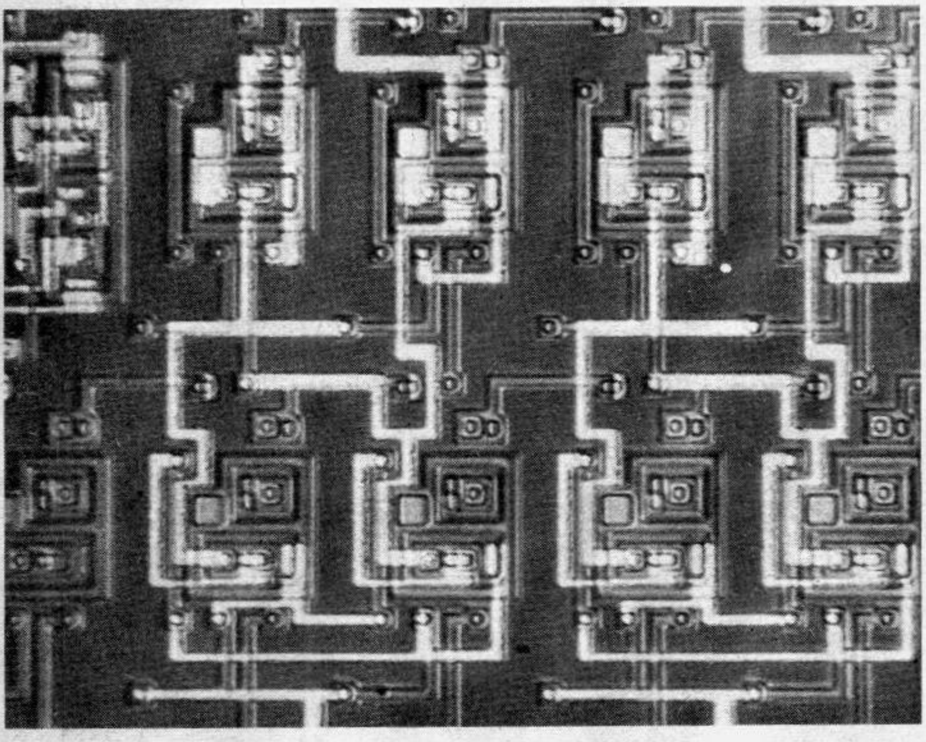
Price erosion

One of the recurrent themes of this series of articles has been the constantly falling price of semiconductors in a society where practically every other commodity was continually costing more. In part this is the result of learning about the processes involved so that production yields have increased. But commercial pressures have also had an important part in forcing prices down. It has been suggested that the combined effect of learning and competition on the price of a device is linearly related to the total quantity which has been manufactured. By examining, for example, the U.S. price of an integrated circuit first produced in 1963 for $30, it can be seen that there has been a price fall of 30% each time the total quantity from the start of production is doubled. Neither does this apply solely to integrated circuits. Germanium diodes, for instance, which were amongst the earliest of semiconductor devices, were originally sold for £1 each. Today their quantity price is 2p and still falling. Although these figures are typical there have been cases where the rate of fall is much greater than this. In early 1969 t.t.l. gates cost 75p, 28p at the end of 1969 and 8p a year later. This was the result of dumping, giving price falls in excess of the learning process, so that prices were so low that U.K. manufacturers could not make and sell devices at these prices either then or in the foreseeable future. Even taking the whole i.c. and transistor scene into account it seems unlikely that more than one or two companies can do much more than break even. Why then continue to compete?

Some have not continued, but the average growth of the electronics industry since World War II has been about 12%



Silicon slice with over 60 chips each about 5mm square. The circuit was commercially available from Ferranti in late 1972 and contains the components of a 200-gate uncommitted logic array and is an example of the use of c.d.i.



Left shows one chip of the 60-chip 200-gate uncommitted logic array produced by the c.d.i. process. Right shows a detail of this chip. The function which the chip performs, analogue or digital or both, is selected by the metalization pattern employed, so that only one mask is involved in changes of function. Supply-line connections within the chip are made, not by the aluminium pattern, but through the silicon.

so that it has doubled in size every six years, whereas the target for national growth has been only 3%. This is one incentive to continue. Whether one believes that there is a national need to retain an independent manufacturing capability depends on the way in which one views the British trading role abroad and at home. However it should be remembered that in 1964, 90% of the U.S. semiconductor industry's output was for military applications, in 1968 the figure was 53% and even in 1970 it was as high as 37%. If our relations with the U.S.A. should ever be strained this source of "raw circuit material" for sophisticated electronics equipment might run more thinly.

It has been said that since the U.K. computer industry is effectively a second-sourcing industry (IBM has over 80% of the world's computer market) the U.K. i.c. market can never be more than a second-sourcing industry. This may well be, but at least it provides an alternative for emergencies, which is what second-sourcing is all about. Further, the developments already described may go some way to loosening the ties between semiconductor and computer manufacture by providing other major outlets for the industry.

Market potential

The accompanying table gives some indication of the areas in which the active device side of the electronics industry has grown. The bare statement, sometimes heard, that the sales of thermionic valves still outstrip those of semiconductors can be seen to fall into that class of lies we call statistics when it is noticed that major valve sales today are of colour television tubes and professional valves such as transmitting valves. Thus the sales of i.cs and transistors are about four times those of receiving valves and are bound to continue to increase as the maintenance market for valve equipment decreases. The apparent setback revealed by the 1971 figures seems rather larger in the table than the true result because ITT, who have approximately 15% of the semiconductor market, have amongst others now withdrawn from V.A.S.C.A. In future, the figures from the Electronic Components Board will include imports from companies such as Motorola and R.C.A. who do not manufacture in the U.K. and General Instrument Microelectronics who are now manufacturing at Glenrothes. Thus the figures for 1972 and subsequent years will not be strictly representative of either British-owned companies or British-made semiconductors.

The annual statistical survey of the electronics industry, published by the U.K. Electronics Economic Development Committee (H.M.S.O., September 1972) dealt more generally with the whole electronics industry employing 446,600 people, a slight drop on 1970's figures. The report which is for the year 1971 said that over 50% of them were employed in S.E. England, the area, understandably, with the greatest unemployment figures for electronics. The survey claimed an 80% growth for the industry in 1971 compared with 21% in 1970. Whilst these figures do not appear to take into account the erosion of the value of the pound, they point out that the reduced rate of growth was most apparent in the capital equipment sector, which includes computer production. The decline of the home demand for computers resulted in a fall of 8% in sales of this type of equipment, thus giving further support for seeking new outlets for microelectronics, such as the colour television market where the growth rate was 72%. That particular level of development obviously cannot continue for long, but it shows what can be achieved in some sectors of the market.

It may well be asked if these sales figures are of interest to readers. Surely whether we are manufacturers of devices or electronic instrumentation, or are industrial or domestic users, or educators, or just have the interest of the success of our country at heart — because our own interests are linked with it — the fortunes of the semiconductor industry matter to us. When we feel the full impact of going into Europe this will be even more vital. The exports from West Germany, for example, are some 70% greater than Britain's and their imports are one third of the total E.E.C. imports. What are our prospects then?

They are in fact very good, especially if recent surveys are any better than former ones. The survey carried out by Mackintosh Consultants for the Department of Trade and Industry in 1971 in the context of the difficulties the industry has been facing, puts the U.K. integrated circuit market at £100 million by 1980 that is an increase of six times over the 1970 figure of £16.5 million quoted in the table, and the m.o.s. share of this at £20 million. Quantum Science Corporation of New York recently forecast a growth rate of 14% per annum for the U.S. market, but both sets of consultants comment that there is not likely to be any diminution of the competitive market forces in the foreseeable future, or to put it more crudely, no one is content to let anyone else have a slice when there is the remotest chance of grabbing it all for himself. And this at a time when it is so difficult to keep unemployment levels down. We must turn a capital-intensive industry, likely to become more so, into one which is certainly intensive of those recently lacking human values.

So there it is: more rough water ahead, more risks for both employer and employee, but high prizes for the far-sighted. It happens, though, that to a great extent we are all in the same boat: let us hope we do not get sea-sick on the way.

Acknowledgement

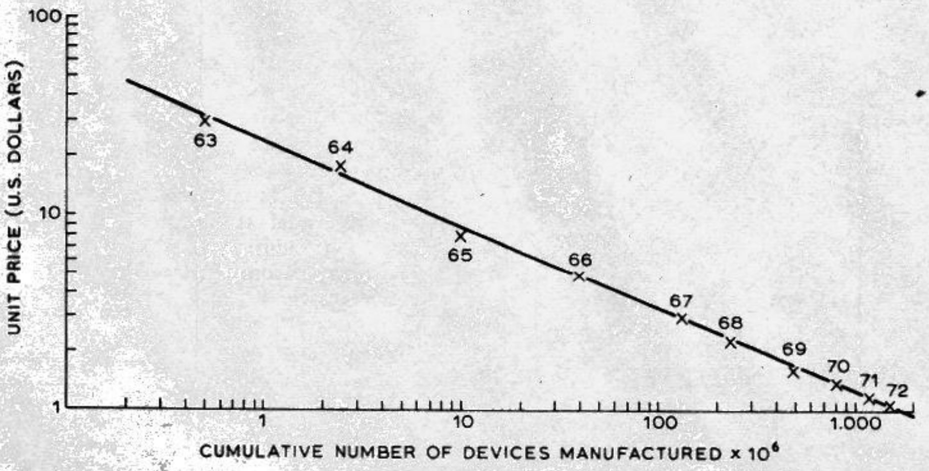
The authors would like to thank their many friends in the semiconductor industry and in Government departments who have racked their files (and their memories) for past detail, and who have been prepared to speak freely about the present and the future.

Sales figures for the U.K. electronics industry.

The figures refer to U.K. based manufacturing plant. Hence T.I. and S.G.S. are included but not Fairchild. The figures were provided by the Electronic Components Board which includes B.V.A., V.A.S.C.A. and R.E.C.M.F. All figures are in millions of pounds sterling and are not corrected for the changing value of the pound.

year	receiving valves	television tubes		professional valves and tubes	discrete semiconductors	integrated circuits
		monochrome	colour			
1955	10.2	10.8	—	—	—	—
1960	14.7	13.9	—	11.6	—	—
1965	13.1	12.8	—	20.1	28.1	—
1970	13.8	17.3	23.0	23.8	46.2	16.5
1971	11.9	16.2	31.7	25.2	34.0*	10.4*

*The figures for 1971 do not include I.T.T.



Relationship between the unit price in U.S. dollars and the cumulative number of devices manufactured between the years 1963 and 1972.