Founded 1925

Incorporated
by Royal Charter 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

Volume 43 No. 1/2 January/February 1973

The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

The First 25 Years of the Transistor

SELDOM, if ever, in the whole history of man's technological achievements can an invention have made such an impact in such a brief time as that of the point-contact transistor at the Bell Telephone Laboratories in 1948. The choice of a period of time at which it is reasonable to step back and review progress and celebrate achievement is inevitably somewhat arbitrary but the 'silver jubilee' does seem to be the most popular. Much however had been achieved in the first decade of the semiconductor era initiated by the epoch-making event we are now celebrating: the award of the Nobel Prize only eight years after the basic invention is a striking testimony to the acknowledged importance of the new technology.

The papers in this special issue of *The Radio and Electronic Engineer* provide a striking account of the almost bewildering sequence of events starting with the original device of 1948, quickly followed by the junction transistor and then by the evolutionary battle of processes which has eventually led to the practical realization of large scale integration. These are the basic stages of the transistor story, but alongside there has been the eager seizing of the opportunities provided by the new techniques. The earlier applications were necessarily limited by the teething troubles of the devices, but as assurance of performance and reliability was gained, the invasion of every field of electronic and radio engineering took place. Several of the papers refer to the successive conquests of apparently impassable barriers of power or frequency. It would be a bold prophet who would nowadays claim any particular performance level or any application to be impossible.

Engineers, with their concern with manufacturing costs, can take much satisfaction from the spectacular reductions in price which continue to be made with every successive development in semiconductor technology. Circuit arrays of a complexity undreamt of a few years ago can be realized at a cost comparable with that of a single component part as fabricated by means of earlier techniques. A significant key to these engineering triumphs has been the application of automation in the broadest sense of the term. Automation and the semiconductor revolution can be considered as roughly comparable in age and it is a moot point as to whether one could exist without the other.

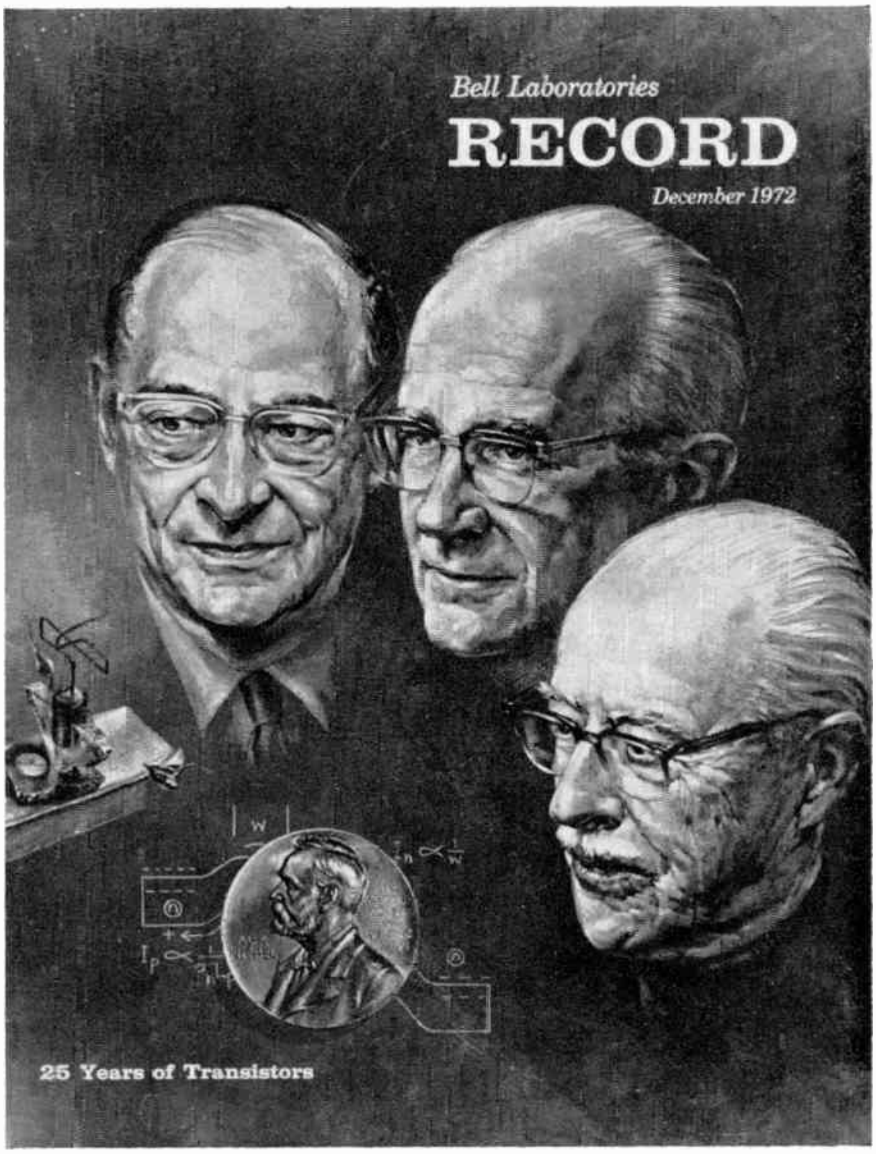
The invention of the transistor is clearly the real reason why electronics has made the great strides which it has done in the past quarter century in revolutionizing the methods of so many other technologies: indeed, in making it possible to do tasks hitherto considered impracticable if not impossible. The immense potentialities for computation and control, typified by the realities of space exploration, can be exploited in almost any field of human activity one chooses to name; my predecessor, Professor Emrys Williams, aptly summed this up in his Presidential Address six years ago, when he suggested that electronics was the biggest 'nosey parker' of all time!

The gathering together of papers on a single, albeit broad, theme to make up this issue creates a precedent for the IERE and is one with which I am happy to be associated. My original invitations to authors to contribute to the issue stated that it was the intention that the papers should review the past, describe the present state of the art, and look a short way into the future, and I stressed also the aim of assessing the influence which the introduction of semiconductors has had on circuits and on systems. Although considerations of time and space available have prevented the inclusion of every design aspect and system application, the group of papers will surely provide a comprehensive and informative report on the first 25 years of the transistor.

A. A. DYSON

'Three men who changed our world'

Under the above headline in the *Bell Laboratories Record* for December 1972, John Bardeen, Walter Brattain and William Shockley, joint winners of the Nobel Prize in Physics in 1956, recall the events leading up to the invention of the transistor in 1948 and comment on its effects. Morgan Sparks, who built the first junction transistor and has subsequently directed semi-conductor work at Bell Laboratories, also reflects on '25 years of transistors', and quotations from contemporary papers by some of the other workers concerned with the more significant of subsequent developments go to make a fitting commemorative issue.



Dr. John Bardeen joined Bell Labs in 1945 as a research physicist and was co-inventor with Dr. Brattain of the point-contact transistor. In 1972 he received his second Nobel Prize in Physics, jointly with L. N. Cooper and J. R. Schrieffer, for a theory of super-conductivity.

He received the B.S. and M.S. degrees in 1928 and 1929 from the University of Wisconsin and the Ph.D. in 1936 from Princeton University. He has received honorary Doctorate degrees from Princeton University and the Universities of Wisconsin, Glasgow and Notre Dame.

He is a Fellow, and former President, of the American Physical Society and a member of the National Academy of Sciences. Dr. Bardeen served on the President's Science Advisory Committee from 1959 to 1962 and currently is Professor of Electrical Engineering and Physics at the University of Illinois.

Dr. William Shockley, presently serving as Executive Consultant to Bell Laboratories, is the inventor of the junction transistor. He joined the Bell Labs staff in 1936 and, after serving with the U.S. Navy in anti-submarine warfare research during World War II, became co-head of the successful solid-state physics research programme.

In 1955 he established Shockley Semi-conductor Laboratory at Palo Alto, California, and continued his work in the transistor field with Clevite Transistor until 1965.

Dr. Shockley became the first Alexander M. Poniatoff Professor of Engineering Science at Stanford University in 1963. He has received honorary Doctorate Degrees from the University of Pennsylvania, Rutgers University and Gustavus Adolphus College. He received the Bachelor of Science degree from the California Institute of Technology in 1932 and the Ph.D. from the Massachusetts Institute of Technology in 1936.

Dr. Walter H. Brattain joined Bell Labs in 1929 as a research physicist in the field of surface properties of solids and subsequently worked in the field of semi-conductor surfaces. These studies resulted ultimately in the invention, with John Bardeen, of the point-contact transistor.

Dr. Brattain received the B.S. degree in Physics and Mathematics in 1924 from Whitman College, the M.A. in 1926 from the University of Oregon, and the Ph.D. in 1929 from the University of Minnesota. He has been awarded honorary Doctor of Science degrees from Portland University, Whitman College, Union College, the University of Minnesota and Gustavus Adolphus College and the L.H.D. from Hartwick College.

During World War II Dr. Brattain was associated with the National Defense Research Committee at Columbia University. He is a member of the National Academy of Sciences.

UDC 621.382

The first decade of transistor development: a personal view

Professor J. J. SPARKES,
B.Sc., Ph.D., C.Eng., M.I.E.E.*

SUMMARY

The development of the junction transistor from its earliest form of a plastics encapsulated, alloy transistor to the final silicon planar device is traced during the period from 1952 to 1962. At the same time some of the significant steps in the development of the understanding of the device are described.

* *Faculty of Technology, The Open University, Walton Hall, Walton, Bletchley, Buckinghamshire.*

1 Introduction

In 1952 I joined an industrial research company and began work studying the behaviour, the properties and the use of transistors. This was four years after the point contact transistor had been discovered and only about three years after the junction transistor had been conceived. By the time I left industry 10 years later a major new technology had been born and had reached maturity. One of the features of working in an area such as electronics is that a period as short as a decade can be significant in the history of a major technological development.

Of necessity the account I am about to give is a personal one. It is a description of what seem to me to be the key developments during the period 1952 to 1962 as the transistor grew from being a low-frequency, unreliable, somewhat mysterious device produced in small quantities one by one, to being a mass-produced, pretty well understood, reliable, high-frequency device used in almost every piece of electronic equipment manufactured.

2 The Beginning

In 1952 in the United Kingdom a few companies had begun to try to manufacture point contact transistors. These devices consisted of a piece of germanium onto the surface of which two point contacts (pointed wires) were placed at a distance apart of perhaps half a millimetre. A capacitor was discharged through one of the point contacts and this contact formed the collector. This device exhibited a current gain greater than one. That is, if the emitter current was changed from, say, 1·0 mA to 1·1 mA, the collector current might change from 0·9 mA to 1·2 mA. Thus $dI_C/dI_E = 3$.

The point contact transistor had been discovered in 1948 by W. Shockley and his associates at the Bell Telephone Laboratories. Shockley had predicted theoretically that by placing a piece of semiconductor, such as that used in the wartime diodes, in an electric field it should be possible to change its resistance and thus produce an electronic amplifier to replace the vacuum tube. Since then this prediction has been borne out in the insulated gate field effect transistor, but at that time the effect was not observed. This failure was (rightly) attributed to some surface effect and in a series of experiments to study the problem a wire was used to inject current into the material and a second one brought up to it to study nearby surface potentials as shown in Fig. 1. Surprisingly any current changes in the input wire were amplified in the output wire, thus giving the current gain of up to around 5 or so already referred to.

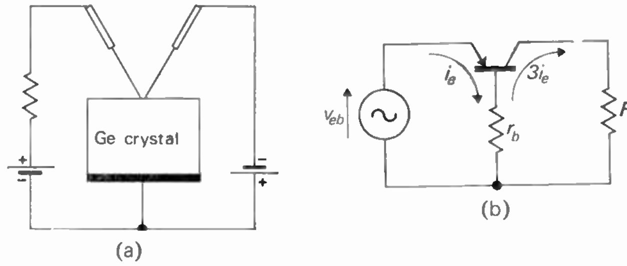


Fig. 1. The point contact transistor.

As can be seen from Fig. 1(b), this current gain of greater than one results in a negative input resistance to the device. Thus if the emitter base voltage is increased, so increasing the current through the emitter junction, the current in the base resistance increases in a direction opposite to that which you would expect from a positive resistance, so that dV_{in}/dI_{in} is negative.

Negative dynamic resistances have always been attractive to circuit engineers since they can lead to very simple switching circuits, such as those used today incorporating thyristors or silicon controlled rectifiers. But the point contact transistor, although exhibiting this effect was also highly unreliable and very noisy. Its noise figure exhibited the usual 1/frequency characteristic but was about 60 dB for a 1 kΩ source at a frequency of 1 kHz. It used to be said of the point contact transistor that if you extrapolated the $1/f$ law characteristic down to a frequency of about 1 cycle per month, the noise voltage it generated, still on the $1/f$ law, was sufficient to blow it up! This, it was said, accounted for the device's high failure rate.

In 1952 junction transistors, with fairly stable current gains of less than unity, were just beginning to become available, and their arrival heralded the demise of the point contact transistor, despite its remarkable negative-resistance properties.

Also in 1952 the Institute of Radio Engineers in America published in their *Proceedings* their first Transistor Issue. This brought together a range of really penetrating papers dealing with this new technology. In particular it described the alloy junction transistor, which for several years dominated the market in Europe.

The story I wish to relate could very well be described as a battle to overcome the deficiencies of the alloy junction transistor, terminating in the production of the silicon planar transistor. So I shall begin by describing the alloy junction transistor in the form in which it was first described by R. R. Law[1] of RCA and by J. S. Saby[2] of General Electric.

Both authors referred to the physics of the device as worked out by Shockley, Sparks and Teal,[3] and by Shockley.[4] The essential features of the device were described as follows:

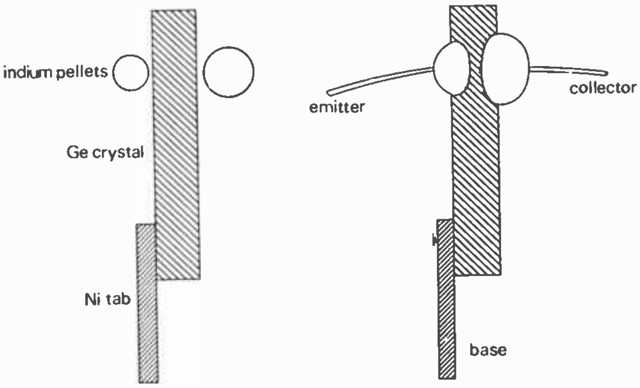
'The junctions are made by alloying indium into opposite faces of the germanium wafer,' and

'For protection the completed unit is embedded in a lightproof, water-resistant plastic.' A cross-section of the transistor is shown in Fig. 2(b).

This device, which could be obtained in limited quantities in this country, had a cut-off frequency around 1 MHz, although exceptional units reached about 3 MHz. The plastic encapsulation led to a fairly rapid deterioration in the performance of these devices, and so far as I know they were never incorporated in any equipment.

They were soon followed in Europe by equivalent devices, the OC11 and OC12. These were encapsulated in black plastic, whereas the RCA ones were white, but this unfortunately resulted in no improvement. The deterioration with time which occurred was a decrease of the current gain and an increase in the cut-off currents of the emitter and collector junction.

Even though the device technology was in its infancy a major step forward in the theoretical understanding of the device was also published in 1952 by J. M. Early[5] in the same issue of the *Proceedings of the IRE*. This paper described the effects of the widening of the collector junction as the collector voltage was varied and from this effect Early was able to deduce accurate equations describing the small signal performance of the device. It led him to propose the T-equivalent circuit of the transistor and he was able to work out the theoretical values of the elements of this equivalent circuit.[6]



(a) Before alloying process. (b) Final cross-section.

Fig. 2. Early alloy junction transistor (not to scale).

Now although RCA, and other firms, were able to improve the technology of the manufacture of these alloy transistors to a point where higher currents could be handled and a cut-off frequency of 10 MHz or so could reliably be achieved,[7] the structure suffered from a basic deficiency. It was manufactured by first producing a thin wafer of about 0·005 cm thick and then placing indium pellets on opposite sides of the wafer from which the alloy junctions were formed as indicated in Fig. 2(a). The final base region between the two alloy junctions upon which the cut-off frequency depends had to be of the order of 0·001 cm or less. But the thickness of the base region between the emitter and collector could not be controlled to any greater accuracy than the thickness of the original wafer. It was the difficulty of controlling this initial thickness and of controlling the depth to which the alloy penetrated which made it impossible to achieve a high yield of devices which had a high cut-off frequency. The next major step in overcoming this deficiency was achieved by Philco with their surface barrier device.

3 The Surface Barrier Transistor

At the time, the surface barrier technique seemed to be a real breakthrough and, in its most advanced form looked as though it might dominate the whole transistor industry. It is a measure of the rate and extent of the progress in transistors that even in this advanced form, the micro-alloy diffused transistor, it is now of historic interest only.

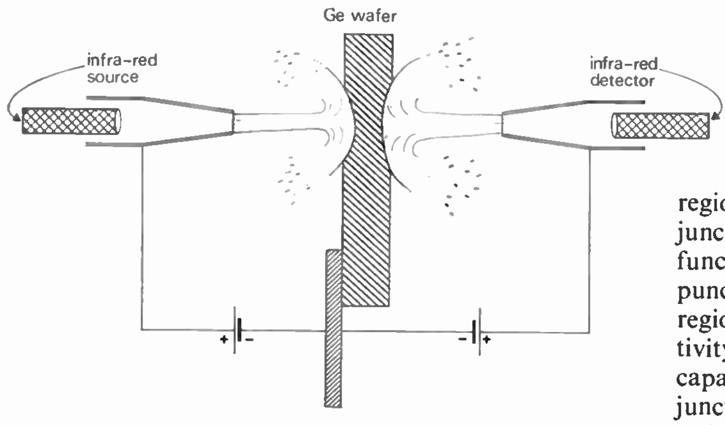


Fig. 3. Surface etching of surface barrier transistors.

The surface barrier transistors were announced by Bradley and others in late 1953.[8] Essentially the technique was as follows. The process started with the same germanium wafer as the alloy process but this time the wafer was mounted between two small jets of electrolyte as shown in Fig. 3. The germanium wafer was held at a positive potential with respect to the nozzles of the jets, so that the electrolyte etched away the surface of the germanium on the opposite sides of the wafer. The ingenious addition to this technique, which changed it from being a minor advance to something of a breakthrough, was the addition of a source of infra-red radiation and a detector to monitor the thickness of the wafer. The infra-red radiation shone through both jets and through the germanium wafer, and as the wafer became thinner the infra-red transmission increased. The wave length of the infra-red radiation was chosen to be on the absorption edge of germanium and so gave a sensitive indication of the thickness of the material. At the correct level of intensity of transmitted light, corresponding to a particular wafer thickness, the polarity of the voltage applied between the wafer and the jets was reversed. At this point etching stopped and metal was plated on the new clean surfaces of the germanium. It was discovered that these surfaces in intimate contact with the germanium were rectifying junctions just like p-n junctions. However it was soon found that a very brief 'micro-alloying' process improved the properties of these junctions considerably and was added to the process before long.

This production technique made it possible reliably to produce base widths much narrower than could be produced by the alloy process. In addition, the technique lent itself to automatic production, and very sophisticated automatic production lines were built to produce these transistors.

Surface barrier transistors had cut-off frequencies in the neighbourhood of 60 MHz and gave great encouragement to the radio industry. However, they had a number of limitations. The narrowness of their base region reduced the maximum operating collector voltage to well below 10 V, owing to the ease with which punch-through occurred. The same widening of the collector transition region which leads to the Early effect can also lead to the widening of the collector transition region right through the base region as far as the emitter junction. When this happens the transistor ceases to function and the effect is called 'punch-through'. Now punch-through can be diminished by doping the base region more heavily, and thus producing a lower resistivity base region. But this increases the collector capacitance, and reduces the breakdown voltage of both junctions. In addition the base resistance (i.e. the resistance of the germanium between the base terminal and the base region) was high in these devices—several hundred ohms. Indeed the frequency limitations arising from the base resistance–collector capacitance time-constant were more severe than those arising from the transit time across the base region. So the surface barrier transistor was only a limited solution.

4 The Graded Base Region

The really important development, due to Kroemer[9] was the introduction of a graded base region. Hitherto transistors had been made with a homogeneous doping density in the base region. Kroemer showed that by grading the doping density from a high density near the emitter junction to a low density near the collector junction as shown in Fig. 4, several benefits accrued. Firstly, this density gradient of donors (in a p-n-p transistor) produced a similar grading of electrons. Such a gradient would, if unconstrained, lead to a diffusion current of electrons from emitter to collector. However, in a p-n-p transistor, the electrons in the base region cannot flow out through the collector junction, so a field is set up in the base region which just balances their tendency to diffuse. But, this field is in the direction which assists the flow of holes across the base

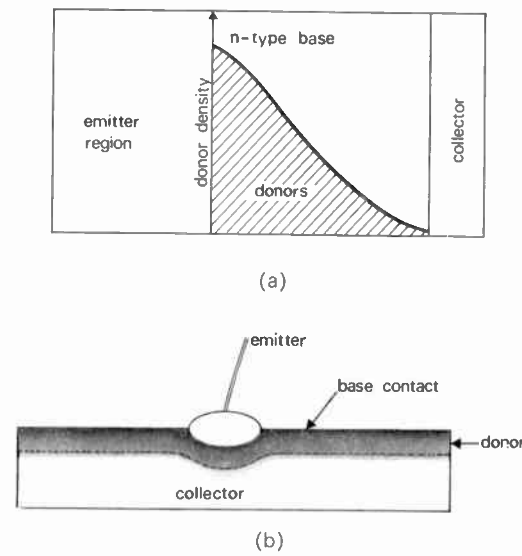


Fig. 4. The graded base region. (a) A graph of density versus distance through the base region; (b) A cross-section indicating donor density and the low resistivity surface of the base region.

region, and therefore facilitates the flow of minority carriers from emitter to collector. Thus the inclusion of a graded base region in a transistor significantly reduces the transit time of minority carriers across it. Increases in cut-off frequency by a factor of three, four or five could readily be achieved by grading the impurity density in the base.

It turns out however that this property of the graded base region is of much less significance than two other properties.

The presence of a high doping density of impurities in the base region next to the emitter junction ensures that there is a low resistivity in the base region close to the emitter as indicated in Fig. 4. Thus a base contact placed close to the emitter contact gives a relatively low base resistance.

The second property is that the problem of punch-through voltage is almost eliminated. The base region is lightly doped near the collector junction, so leading to a low collector capacitance, a high collector breakdown voltage, and also leading to a significant widening of the collector transition region as the reverse voltage on the collector is increased. However with this structure the wider the collector transition region becomes, the more highly doped becomes the base region into which the transition region penetrates. Indeed the doping density of the base region becomes so great near the emitter that punch-through does not occur. Thus the combination of high punch-through voltage with low collector capacitance, which cannot be achieved with a homogeneous base region, was an automatic result of the graded base region. Not surprisingly ever since Kroemer's invention almost all new designs have included graded base regions.

Now when the surface barrier technique was used with germanium wafers into which a gradient of impurities had been diffused, the devices known as the micro-alloy diffused transistors[10] were produced, and these appeared to combine almost all the advantages which the theory relating to low power devices could name. The theoretical problems seem to have been solved. However, the original range of problems began to become of more importance again, namely the reliability, reproducibility, and stability with time of the parameter values. In addition, although the device could be produced automatically, it could not be mass produced. That is to say each device was made individually, they were not made in thousands at a time and therefore could never exhibit the economies of production characteristic of mass-produced devices. However for a few years the micro-alloy-diffused transistor, the alloy-diffused transistor,[11] which in essence was the alloy transistor made using a graded impurity wafer, and the double-diffused transistor known as the mesa transistor,[12] dominated the high frequency transistor market. Indeed it was a development of the mesa transistor which eventually won the day, as I shall describe after the next section.

5 Theoretical Developments

Apart from Kroemer's demonstration that a graded base region would reduce the transit time of carriers across the base region of a transistor, in 1954 there were made, in my view, three principal theoretical contributions to the understanding of transistors. Today two of them still play a key role in our analysis and understanding of the devices.

The first is the discovery of the hybrid-pi equivalent circuit by L. J. Giacoletto.[13] Before his exhaustive study published in *RCA Review* most engineers used a T equivalent circuit of one kind or another. Usually they had been derived theoretically from physical considerations concerning the structure of the device and, of course, approximations in the theoretical analysis of different kinds led to different simplified equivalent circuits. Giacoletto pointed out that physical understanding can lead to a number of possible equivalent circuits. He determined the parameters of various equivalent circuits experimentally over a wide frequency range. From these studies emerged the fact that the hybrid-pi equivalent circuit, shown in Fig. 5, gave a simpler representation of a transistor's performance than any of the others he tried. In addition it only contained one generator whereas the best T-equivalent circuits always contained a feed-forward generator as well as a feed-back generator.

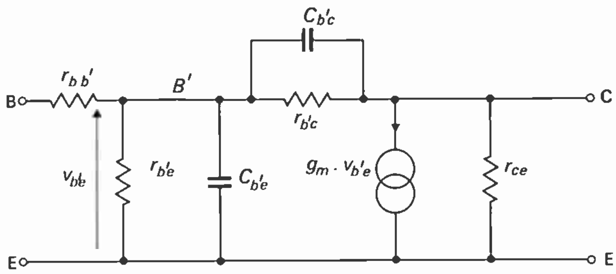


Fig. 5. The hybrid-pi equivalent circuit.

Perhaps the most important feature of the hybrid-pi equivalent circuit is that it can be used at various levels of accuracy. For the simplest intuitive design procedures it is possible to strip away all the equivalent circuit elements shown in Fig. 5 and keep only the current generator. This generator is characterized by the mutual conductance, g_m which is almost 40 mA/V at a collector current of 1 mA for all well-behaved transistors. This value of mutual conductance does not depend to any great extent on any variations in the design of the device provided all parts of it are behaving properly. It is a fundamental quantity in transistors.

To achieve greater accuracy in the equivalent circuit, especially as regards input and output resistance, $r_{b'e}$ and r_{ce} can be added. For higher frequencies or for higher currents then $r_{bb'}$, $C_{b'e}$ and $C_{b'e}$ can be added. And finally to obtain even better representation $r_{b'c}$ can be added. Furthermore each component of the equivalent circuit represents a fairly clear-cut physical process within the device, so it is possible by understanding the structure of the device, as well as by understanding the circuit in which the device is to be used, to judge which elements of the equivalent circuit should be included in any particular design problem.

A second important contribution was made by J. J. Ebers and J. L. Moll.[14] They derived for the first time a pair of equations representing the d.c. characteristics of a transistor over the whole range of possible bias conditions. Their equations are:

$$I_E = I_{ES}\left[\exp\left(\frac{q\Phi_E}{kT}\right)-1\right] - \alpha_R I_{CS}\left[\exp\left(\frac{q\Phi_C}{kT}\right)-1\right] \quad (1)$$

$$I_C = I_{CS}\left[\exp\left(\frac{q\Phi_C}{kT}\right)-1\right] - \alpha_N I_{ES}\left[\exp\left(\frac{q\Phi_E}{kT}\right)-1\right] \quad (2)$$

Φ_E and Φ_C are respectively the emitter-base and collector-base junction voltages.

An alternative form of these equations which is much easier to understand is shown in equations (3) and (4):

$$I_E = I_{EBO}\left[\exp\left(\frac{q\Phi_E}{kT}\right)-1\right] - \alpha_R I_C \quad (3)$$

$$I_C = I_{CBO}\left[\exp\left(\frac{q\Phi_C}{kT}\right)-1\right] - \alpha_N I_E \quad (4)$$

These equations can be interpreted as follows. Equation (3) says that the emitter current I_E is the normal diode current through the emitter p-n junction, represented by the usual characteristic equation for a diode, together with a proportion, namely α_R, of the current, I_C, flowing in the collector junction. Similarly, by eqn. (4), the collector current is the collector diode current minus a proportion, α_N of the emitter current. Thus the equations simply express the interaction between p-n junctions which is to be expected. Like all important contributions, once they have been made, it is difficult to understand why they were not made earlier.

These equations, as Ebers and Moll described in their paper, do not express all the physical effects within a transistor. In particular the Early effect, referred to earlier, in which a variation of the voltage across the junction affects the width of the transition region of that junction, is not included. This means that the factors α_F and α_R are regarded in these equations as independent of the voltages applied to the junctions and lead to a slope resistance of the output characteristics of a device which is quite wrong. A further inadequacy in the equations, not appreciated till quite recently, is that they do not express the fact that in silicon junctions a significant proportion of the base current arises from recombination within the p-n junction, to which the characteristic diode equation does not apply. Thus the base current in a silicon transistor has a different index to the exponential term in the diode equation than does the collector current. But this takes me outside the first decade with which this paper is concerned.

The third development at this time was the first treatment by Moll[15] of the transient response of transistors. At that time many engineers were attempting to use the transistor as a switching device, because it was known to have extremely good switching characteristics. The Ebers and Moll equations were able to predict very low saturation voltages under certain circumstances, and these were borne out in practice. But these d.c. equations were unable to describe the switching speed of a device, and at that time most circuit designers were feeling their way very uncertainly. In December 1954 Moll presented a whole range of equations relating switching speeds to circuit conditions and configurations. He showed for the first time that the inverse operation of the device (that is, its behaviour when the collector is used as an emitter, and vice versa), affected the transient response. The analysis however turned out to be only a half-way stage towards a full understanding of the switching properties of transistors.

A second step in understanding the switching properties of transistors came in 1957 when rather than express the large signal transient response in terms of small signal parameters, Beaufoy and Sparkes[16, 17] showed that a simpler and more complete description of the behaviour was to relate the output current, I_C, of the transient to the base charge, Q_B, of the device. In other words as far as the large signal transient response is concerned the most precise transfer parameter describing the action of a transistor is a time-constant I_C/Q_B, the 'collector time factor'. Thus if a circuit designer requires a given rate of change of collector current, then, from a knowledge of the collector time factor, he could calculate the rate at which it was necessary to supply charge to the base of the transistor. With this knowledge he could then design his input circuitry. In 1957 equipment was being designed in large quantities using alloy junction transistors of various kinds and the charge required in the base of these transistors was the dominant factor in determining their transient response.

Only a few years later however, after the introduction of much higher frequency switching transistors, this particular charge ceased to be the dominant factor in determining switching speed. Nevertheless the general philosophy of regarding input charge as a basic design parameter has remained in the design of many high-speed switching circuits.

The particular advantages of transistor analysis in terms of charge control are, first, that although the output current can be expressed in terms of input current, input voltage or input charge, the inherent delay in the device between a given input and the appearance of the corresponding output is least when it is related to charge. The second is that it leads directly to simple theoretical expressions for the ultimate possible switching speeds of solid-state devices.[18]

6 The Development of the Planar Transistor

As already mentioned, the mesa transistor was one of the high frequency devices which made use of the graded base configuration invented by Kroemer. The device was manufactured as follows (see Fig. 6).

The process began with a wafer of germanium or silicon whose resistivity was that required for the collector region of the transistor. Let us consider an n-p-n transistor, so the starting wafer is of n-type material. The next step was first to diffuse in some acceptors and so convert the outside layer of the wafer into p-type material. This was then followed with a much heavier doping with donors, but also a more rapid one, so that a final thin outside layer of heavily doped n-type material

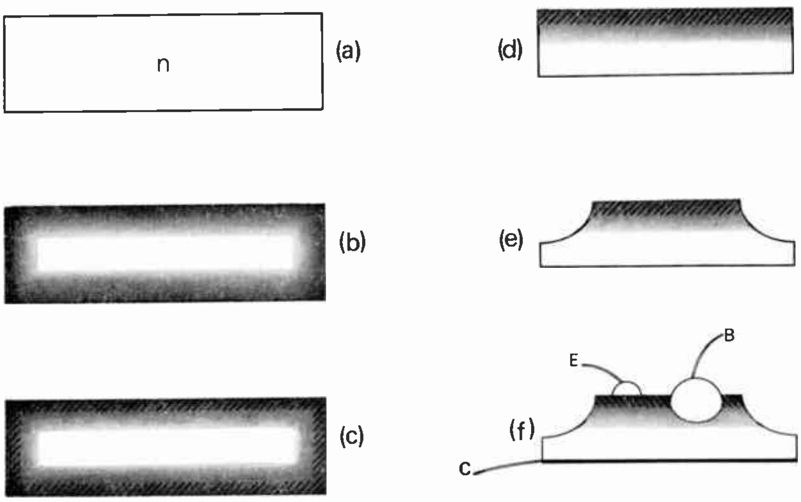


Fig. 6. Manufacture of the mesa transistor:
(a) the n-type starting wafer;
(b) after a p-type diffusion;
(c) after an n-type diffusion;
(d) removal of unwanted diffused layers;
(e) the 'mesa' produced by etching;
(f) electrical contacts are made.

was produced as shown in Fig. 6(c). In this way the whole of the outside surface of the wafer became an n-p-n sandwich. After removing these layers from three of the surfaces the problem then was simply to divide up the wafer into small areas and make each one into a transistor. However, when the wafer was divided up into small dice, as small as could conveniently be handled, the p-n junction areas of both emitter and collector were too large to give high frequency performance. That is, the collector capacitance, as a result of its large area, was higher than was needed. The final process therefore was to etch away some of the upper surface and leave only a small platform, or 'mesa', standing up on the broad base of the collector region, as shown in Fig. 6(e).

Problems still remained: in particular the attachment of the base lead was awkward, and a number of processes were developed to make this easier. For example, the alloy-diffused process, giving p-n-p transistors, did away with one diffusion and used an alloy process to attach the emitter. Included in the emitter material was a fast diffusing n-type impurity, so that during alloying a diffusion into the base region occurred, thus causing the required grading of acceptors in the p-type base region.[11]

There were however, two more serious difficulties with the process.

First, there was no satisfactory choice of resistivity for the collector regions which met all the requirements placed upon it. The collector region had to be sufficiently thick to ensure that mechanical handling of the wafer did not break it. This limit had been found many years earlier and was still about 0·005 cm. Since the base region was now less than 0·0005 cm it is clear that the collector region was much thicker than was necessary for the electrical performance of the transistor. To obtain a low collector capacitance and a high breakdown voltage it was desirable that this collector region should have a high resistivity. However, a high resistivity produced a significant and undesirable resistance between the collector junction and the collector contact.

The important solution to this problem was the achievement of the epitaxial process[19] for depositing a thin layer of high resistivity material on top of a thick substrate of low resistivity material. Diffusion is a useful process for creating layers of low resistivity on high resistivity substrates, but it does not work satisfactorily the other way round. The epitaxial process is a vapour deposition of high purity semiconductor on a lower purity substrate which nevertheless continues the crystal structure of the substrate. Now, if the transistor process is started with a low resistivity substrate upon which a high resistivity layer has been epitaxially deposited, the diffusion processes can take place in the epitaxial layer only, and so achieve the required low capacitance and high breakdown voltage. But the necessarily thick substrate does not now have a high resistance.

The second and most important of all developments was the protection by the planar process of the vulnerable edges to the p-n junctions, where they are exposed to the environment. As can be seen from any of the cross-sections shown in this paper, the p-n junction comes to the surface of the semi-conductor material and is exposed to the environment. All transistors are encapsulated to ensure protection of the semi-conductor surface, but even so it had been known for a long time that no encapsulation was sufficiently stable to ensure that the surfaces did not change with time. For example, traces of moisture would evaporate or condense with change of temperature. The result was that the current gain and the leakage currents of the transistor were subject to variations. It was known for example that the inclusion of water vapour would increase the current gain, and some manufacturers deliberately included traces of water vapour within the encapsulation to improve the transistor current gain. But the inevitable result was that the stability of the current gain deteriorated.

The planar process, which is now virtually the only process used for silicon devices, was first described in Europe at a conference in Paris in 1961. Victor Grinich[20] of the Fairchild Corporation presented graphs of the change with time of current gain, base-emitter voltage and cut-off current of planar transistors which were so much better than anyone had seen before that it was quite obvious that if they were genuine a real break-through had been achieved. After several hours' discussion with Grinich it became clear to me that the planar process was the process of the future. It was an unpalatable conclusion since, just at that time, many companies had recently invested large sums of money in the double-diffused, the alloy-diffused or the micro-alloy-diffused process, with the hope of achieving a clear production run of a few years.

The significant achievement of the planar process (Fig. 7) was the deposition of an oxide layer over the surface of the silicon wafer before the diffusion process occurred. Holes were then etched in the silicon oxide and the acceptors for the p-type base region diffused in as shown in Fig. 7(b). As can be seen the exposed edge of the p-n junction is formed underneath the previously deposited oxide layer. This provides the protection the p-n junction needs. The emitter junction is formed similarly as shown in Fig. 7(c). Evidently, not only are the junctions protected, but the problem of connecting the base lead is satisfactorily dealt with too. With the expitaxial process used to prepare the substrate, the structure can be adapted to almost any design required. The surfaces are in fact so well protected that the devices are quite stable even without encapsulation.

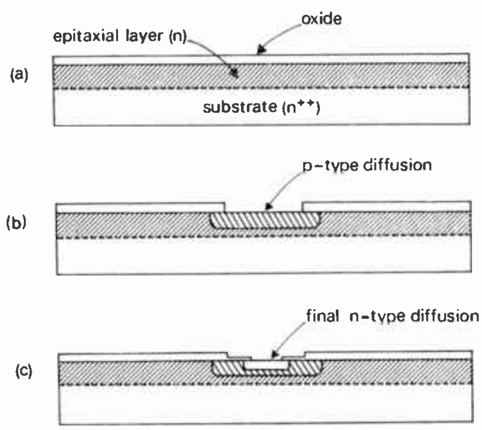


Fig. 7. The planar process. (a) The starting wafer with thin epitaxial layer (not to scale) and oxide coating; (b) The p-type diffusion through a hole etched in the oxide; (c) The emitter, n^+ diffusion through a smaller hole in the regrown oxide layer.

Thus at the end of the decade from 1952 to 1962, the semiconductor industry was at last provided with the technology which was basically sound and which, as many pointed out, held great promise for integrated circuit construction in the near future.

7 References

1. Law, R. R. *et al.*, 'A developmental germanium p-n-p junction transistor', *Proc. Inst. Radio Engrs*, **40**, pp. 1352–57, November 1952.
2. Saby, J. S., 'Fused impurity p-n-p junction transistors', *Proc. IRE*, **40**, pp. 1358–64, November 1952.
3. Shockley, W., Sparks, M., and Teal, G. K., 'P-N junction transistors', *Phys. Rev.*, **83**, pp. 151–62, 15th July 1951.
4. Shockley, W., 'The theory of p-n junctions and p-n junction transistors', *Bell Syst. Tech. J.*, **28**, pp. 435–89, July 1949.
5. Early, J. M., 'Effects of space charge layer widening in junction transistors', *Proc. IRE*, **40**, pp. 1401–6, November 1952.
6. Early, J. M., 'Design theory of junction transistors', *Bell Syst. Tech. J.*, **32**, pp. 1271–1312, November 1952.
7. Armstrong, L. D. *et al.*, 'p-n-p transistors using high-emitter-efficiency alloy materials', in 'Transistors I', pp. 144–52 (RCA Laboratories, Princeton N.J., 1956).
8. Bradley, W. E., 'Principles of the surface barrier transistor', *Proc. IRE*, **41**, pp. 1702–6, December 1953.
9. Kroemer, H., 'Der Drift Transistor', *Naturwissenschaften*, **40**, pp. 578–9, November 1952, 'The drift transistor' in 'Transistors I', pp. 202–20 (RCA Laboratories, Princeton, N.J., 1956).
10. Thornton, C. G. and Angell, J. B., 'The technology of micro-alloy diffused transistors', *Proc. IRE*, **46**, pp. 1166–77, June 1958.
11. Jochens, P. J. W. *et al.*, 'Construction and electrical properties of a germanium alloy-diffused transistor', *Proc. IRE*, **46**, pp. 1161–5, June 1958.
12. Lee, C. A., 'A high frequency diffused base germanium transistor', *Bell Syst. Tech. J.*, **35**, pp. 23–24, January 1956.
13. Giacolleto, L. J., 'Study of p-n-p alloy junction transistors from d.c. through medium frequencies', *RCA Rev.*, **15**, pp. 506–62, December 1956.
14. Ebers, J. J. and Moll, J. L., 'Large signal behaviour of junction transistors', *Proc. IRE*, **42**, pp. 1761–72, December 1954.
15. Moll, J. L., 'Large signal transient response of junction transistors', *Proc. IRE*, **42**, pp. 1773–84, December 1954.
16. Beaufoy, R. and Sparkes, J. J., 'The junction transistor as a charge-controlled device', *ATE J.*, **13**, No. 4, pp. 320–7, 1957.
17. Sparkes, J. J., 'A study of the charge control parameters of transistors', *Proc. IRE*, **48**, pp. 1696–1705, October 1960.
18. Johnson, E. O. and Rose, A., 'Simple general analysis of amplifier devices with emitter, control and collector functions', *Proc. IRE*, **47**, pp. 407–18, March 1959.
19. Theuerer, H. C., 'Epitaxial silicon films by the hydrogen reduction of $SiCl_4$', *J. Electrochem. Soc.*, **108**, p. 649, 1961.
20. Grinich, V. H. and Hoerni, J. A., 'The planar transistor family', 'Colloque International sur les dispositifs à semiconductors', Paris, February 1961.
    See also one earlier presentation:
    Hoerni, J. A., 'Planar silicon transistors and diodes'. Paper presented at the 1960 Electron Devices Meeting in Washington D.C. (October 1960).
21. An excellent review article up to, but not including, planar technology is:
    Anderson, A. E., 'Transistor technology evolution', *Western Electric Engineer*, July 1959, October 1960, January 1960.

Manuscript received by the Institution on 27th September 1972. (Paper No. 1490/CC154.)

The pre-history of the transistor

Professor W. Gosling,
B.Sc., C.Eng., F.I.E.E., F.I.E.R.E.*

Although the invention of the transistor, as we know it today, is rightly ascribed to William Shockley and his co-workers, it is not generally known that this was a re-invention of a device which had, in fact, been invented several times before. It is virtually certain, however, that the various inventors of the transistor were all quite unaware of each other's work.

The story of semiconductor devices begins in 1873 when Braun published his observations of the rectifying properties of lead sulphide and iron sulphide crystals, and the first applications of semiconductors were in the copper oxide (and later selenium) rectifiers, using metal-semiconductor junctions. It was, however, to Julius Lillienfeld that the first invention of the transistor must be ascribed.

In 1925 Lillienfeld applied for a Canadian patent for what would now be described as a junction f.e.t. The semiconductor used was polycrystalline copper sulphide, and the gate junction was formed between this material and the edge of a thin aluminium foil. We do not know for certain whether Lillienfeld actually built the devices he described, but it seems possible that he did, because he went on to patent alternative transistor designs. A patent filed in 1927 describes a bipolar transistor of semiconductor–metal–semiconductor construction. The importance of a very thin base layer was fully appreciated. In 1928 Lillienfeld patented the insulated gate f.e.t. This was an interesting structure using aluminium foil as the gate electrode, anodized to give the required insulation, over which copper sulphide was subsequently deposited. It was thus a true m.o.s. device. Again it is not known whether the device was constructed, but details of the design, such as the introduction of a groove in the semiconductor to reduce the channel width, are suggestive of practical experience.

Thus by 1928 all three types of transistor with which we are now familiar had been patented by Julius Lillienfeld. He was of course before his time: semiconductors were not well enough understood to make his devices practicable. Indeed, Bloch's theory of wave functions in crystalline solids was only just being formulated. The industrial climate also turned distinctly sour in 1929, and no more was heard of Lillienfeld's work.

In 1930 H. C. Weber, of the Industrial Development Corporation, Salem, Mass., applied for a U.S. patent for a device using a Cu-CuS-CuO-CuS-Cu sandwich. Imbedded in the CuO layer is a fine spiral of wire intended for use as a control electrode. No practical results or characteristics are quoted.

The well-known German engineer O. Heil (inventor of the Heil tube) filed a British patent in 1934 for a transistor based on the copper oxide rectifier but also, perhaps more significantly, in 1936 Holst and van Geel, of the Philips organization, filed a patent in which a control electrode was sandwiched within a selenium–silver junction. This was probably a kind of bipolar transistor, although its operation is obscure. van Geel persisted with his work, despite war circumstances, and took out further patents in 1943 and 1945, all the designs being characterized by a multi-layer structure (in one case selenium oxide–copper iodide–selenium oxide) forming a bipolar device, but with the curious feature that a very thin (5 μm) layer of insulator (polystyrene) was introduced between the semiconductor layers. A somewhat analogous structure was also proposed by Glaser, Koch and Voigt in their patent of 1939.

All these developments came to nothing: science and industry were not ready to use them. The transistor remained unknown until the work at Bell Laboratories in the late '40s gave it a more timely rebirth. However, one cannot help admiring the originality of Julius Lillienfeld, who invented all three major types of transistor before 1928, and wondering about those designs by Van Geel, which incorporated an ultra-thin insulating organic layer through which the current passed. There is now considerable interest in organic semiconductors: could this early work be a pointer to the future?

* *Department of Electrical and Electronic Engineering, University College of Swansea, Singleton Park, Swansea, SA2 8PP.*

Reference

Gosling, W., Townsend, W. G. and Watson, J., 'Field Effect Electronics', pp. 1–11 (Butterworths, London, 1971)'

UDC 621.382

Semiconductor device developments in the 1960s

A. A. SHEPHERD,
M.Sc., Ph.D., F.Inst.P. *

SUMMARY

This paper illustrates the progress of semiconductor device development during the 1960s by reference to various families of active devices (excluding opto-devices).

The reasons for the change during that period from germanium to silicon technology are discussed, and the enormous influence of the silicon planar technique on device design and manufacturing methods is shown. The increasing role of the silicon integrated circuit is described, and the reasons why the integrated circuit, both bipolar and m.o.s. will continue to form the basis of future equipment design are explained.

* *Ferranti Ltd., Electronic Components Division, Gem Mill, Chadderton, Oldham, Lancashire.*

1 Introduction

The development of semiconductor devices during the 1960s was dominated by the improvements during that period in the technology of silicon. The recognition by the device user of the enormous potentialities of silicon technology, particularly after the introduction of the planar technique, led to unprecedented pressure for the extension of the use of solid-state devices to new fields of activity. This resulted in turn in an enormous improvement in performance of devices, for example in frequency and power handling capacity. At the same time, the semiconductor industry was improving its production methods to reduce the cost of devices, allowing them to be used in an ever increasing range of equipments, many of which, e.g. the automotive industry, had not previously employed electronic devices.

Many sectors of the device industry expanded sixfold and more during the decade in monetary turnover. At the same time, the cost of individual devices to the customer was reduced enormously due to improved technology. In 1960, the average silicon transistor cost about £1; in 1970, an integrated circuit incorporating over 1,000 such devices, together with diodes, resistors and interconnexions could be obtained for less than £5.

It seems unlikely that industrial growth on the scale exhibited by the semiconductor device industry during the period from 1960 to 1970, together with the accompanying dramatic improvements in production technology, has ever been equalled.

In surveying the progress of device development in this paper, it is not possible to cover the whole field, but some of the more important areas of activity are chosen as illustrations of the changes taking place in the last decade.

2 Small-signal Diodes

The swing from germanium to silicon technology was already becoming evident in the field of small-signal diodes in 1960. Although in the 1950s the great majority of small signal diodes in use were germanium point contact devices, these units were restricted in current handling capacity to a few milliamperes, and because of this tended to be used in applications such as r.f. detectors where their performance was adequate. These devices were also widely used in computer circuitry, often as the input elements to d.t.l. gates.

In spite of their wide usage, the disadvantages of low inverse voltage rating, high reverse current, and poor performance at high temperature made it attractive to the user to turn to the newer silicon units which were derived from the initial work of Pearson and Sawyer at Bell Telephone Laboratories as early as 1952.[1]

The first types to become commercially available were those based on the alloy process in which the junction was formed by alloying an aluminium wire or pellet into an n-type silicon chip, and the typical I/V characteristic of such a unit is shown in Fig. 1, together with a typical germanium point contact diode characteristic. In this device, the principle of conductivity modulation was used in which the base layer was made sufficiently thin for its series resistance to be swamped by injected carriers. The advantages of the device over the germanium unit were higher current handling capacity, higher voltage rating, lower inverse current, and higher temperature rating. This last property was a vital factor for use in military equipments and these diodes were in great demand for such uses.

The new silicon diodes had some disadvantages, however, and one of these was minority carrier storage which led to reverse current spikes when the diodes were used for high frequency rectification. In this respect they were inferior to the germanium point contact units which had a very small active area and low minority carrier lieftime in the vicinity of the junction.

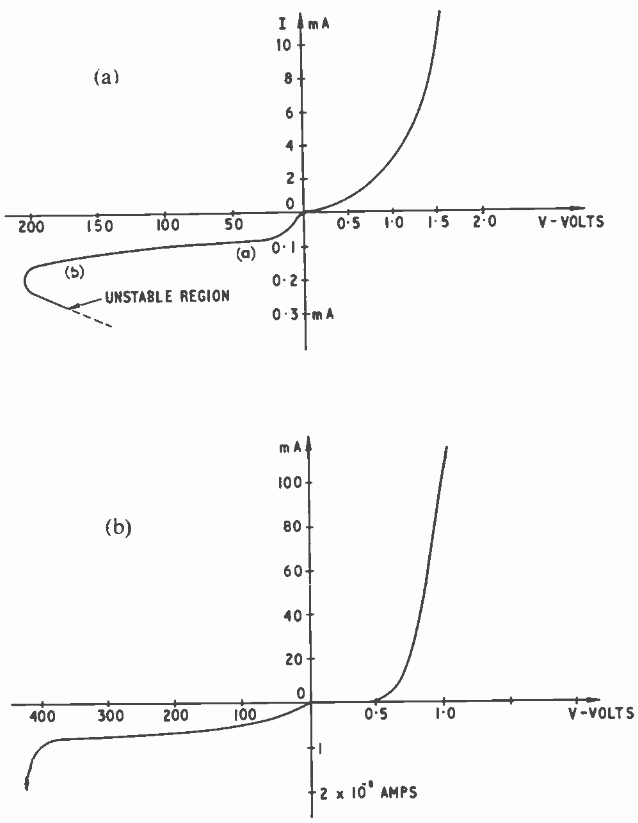


Fig. 1. Comparison of germanium point contact diode characteristic (a) with that of an early silicon diode (b).

The enormous improvement in reverse characteristic can clearly be seen, together with the increased forward current handling capacity of the silicon diode.

This meant that, while the silicon devices were preferred for most uses in the early 1960s, a lot of effort was being put into the problem of making diodes with low minority carrier lifetime, suitable for high frequency operation and fast switching in digital computers, the circuit operating speeds of which were increasing very fast at that time. These aims were realized by 'quenching', or rapidly cooling the diodes after the alloying cycle, thus producing structural defects which reduced the minority carrier lifetime at the expense of an increase in reverse current. It was some time later before the role of gold in the reduction of minority carrier lifetime in silicon was appreciated. Many of these earlier silicon diodes were made by alloying aluminium wire to the silicon to form the junction, and antimony-doped gold wire for the ohmic contact. This gave a structure which was easy to mount in a single-ended package (see Fig. 2(a)). Although equipment makers were, by the early 1960s, accustomed to mounting transistors in similar packages on printed circuit boards, they never fully accepted the single-ended diode package, presumably because the previous germanium diodes had been made with a double-ended construction.

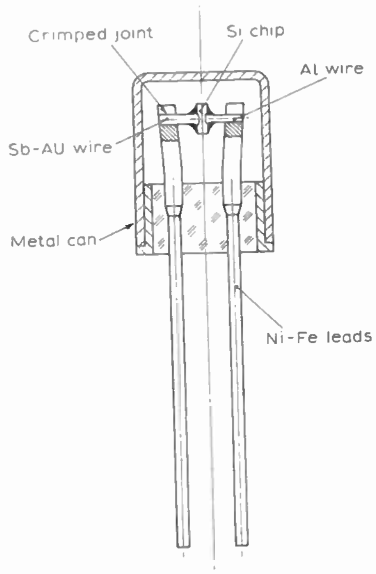
In the early 1960s the customer preference for the double-ended form of encapsulation, together with the fact that this package was cheaper to manufacture, led to the emergence of silicon diodes in this form. Alloyed units were difficult to encapsulate in this way; it was necessary to make the rectifying element in the form of a thin disk of silicon with an aluminium button alloyed to it, and mount this into one end of the glass package first, subsequently contacting the aluminium button with a spring contact as the other end of the package was closed. Although this type of silicon diode was marketed for some years, the spring contact onto aluminium gave poor performance under shock and vibration conditions.

In the meantime, solid-state diffusion as a method of junction formation had become more widely used. Early diffused junction rectifiers were reported by Prince in 1956.[2] Although the main aim of this work was to make power rectifiers without the expansion matching problems involved in alloying large area aluminium disks to silicon, Prince recognized the potential of the diffused junction in replacing alloyed diodes in small-signal applications. He used the diffusion technique to make double-ended glass silicon diodes with current ratings of less than 1 A, using a silicon chip 0·030 in square, encapsulated in a double-ended glass package.

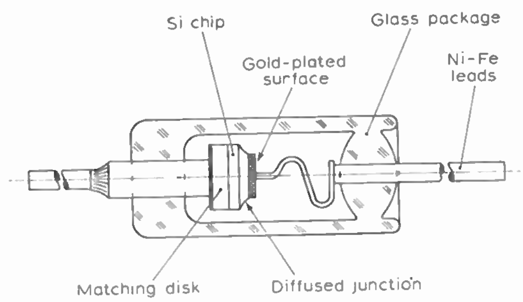
In the period between 1960 and 1965 there was an enormous growth in the use of diodes of this type, notably the 1N645 and 1N649 in the USA, corresponding to the UK types CV7045 and CV7046. These diodes had current ratings of 400 mA at 25°C and peak inverse voltages of 200 V and 400 V respectively. They were capable of operating up to 200°C and were probably the most commonly used diode in military equipments for several years. The contacting system still employed a spring on one side of the chip, but the spring usually made contact to gold plated silicon and was very reliable and rugged (Fig. 2(b)).

A modified version of this device with the junction area reduced by mesa etching, and with gold doping employed to reduce the minority carrier lifetime was used for fast switching applications as the 1N914 series. These diodes exhibited recovery times less than 5 ns.

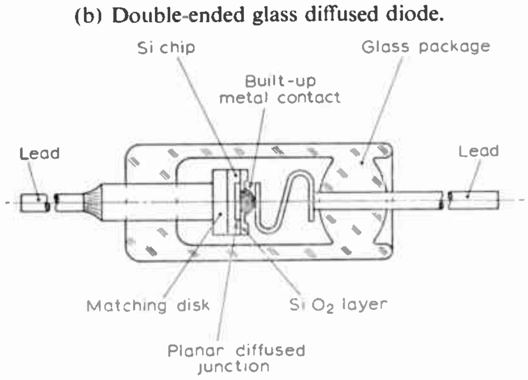
The advent of planar technology in 1962 allowed two major changes in the construction of the signal diode. Firstly, the junction area could be defined by oxide masking before the junction was diffused, and secondly, junction passivation could be achieved, sometimes by the addition of extra passivating oxide glasses, which would stand higher temperature during encapsulation.



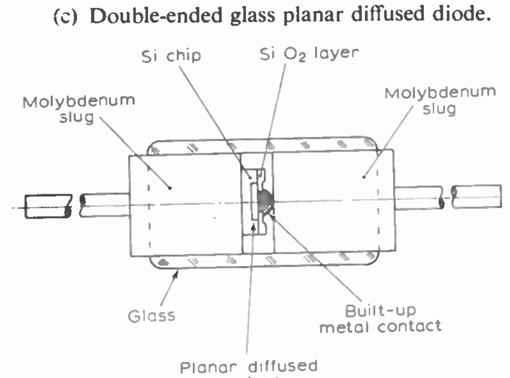
(a) Early single-ended diode.



(b) Double-ended glass diffused diode.



(c) Double-ended glass planar diffused diode.



(d) Double-ended 'solid glass' planar diode.

Fig. 2. Developments in silicon diode manufacture during the 1960s.

A typical planar diode structure is shown in Fig. 2(c), while the 'solid glass' type of encapsulation which came into use in the second half of the decade[3] is shown in Fig. 2(d). Automatic assembly methods have made this type of diode extremely cheap to produce, and the elimination of spring contacts produces a very high degree of ruggedness.

3 Power Rectifiers and Thyristors

In 1960 a substantial proportion of the solid-state power rectifier field was still served by germanium alloy rectifier units of the type originally described by Rouault and Hall.[4] These devices were adequate for most applications up to about 10 A, using conduction cooling. The improvements in germanium material preparation brought about by zone refining techniques allowed these devices to be rated at repetitive peak inverse voltages up to 600 V.

At higher currents, however, it was apparent that in spite of the lower forward voltage drop of germanium units, the power dissipation in the rectifier cell was a limitation, and the current rating was dependent on cooling the rectifier sufficiently well to limit the junction temperature to 75°C. In spite of the fact that individual water-cooled germanium rectifiers could be designed to handle 200 A at a peak inverse voltage of 600 V, the additional complication of water cooling was undesirable, and the rapid improvement taking place in silicon technology at that time made it natural for the power industry to move to silicon rectifiers, with convection or forced air cooling.

Silicon power rectifiers were made in the early 1960s by both alloying and diffusion techniques. The diffused units, developed from the type originally described by Prince in 1956[2] had the advantage that the rectifying junction was within the silicon wafer itself and presented no fundamental expansion matching problem, as did the aluminium/silicon alloy unit. These units often exhibited somewhat inferior inverse characteristics associated with the less perfect junction. However, this advantage was largely overcome by 1956, when 'sandwich' type alloyed structures were introduced, containing molybdenum disks adjacent to the aluminium/silicon element.[5]

The major drawbacks which had to be overcome in introducing silicon high-current rectifiers into the electrical power industry were a sensitivity of these devices to current and voltage overloads. Equipment makers who had in the past used vacuum or mercury arc rectifiers had to embrace a new philosophy in recognizing the low thermal capacity of the silicon rectifier, and the catastrophic results of inverse voltage spikes, even of very short duration, which would not have affected the devices previously used.

The 1960s, therefore, were years in which the development of silicon power rectifiers continued in the direction of higher current and voltage ratings,[6] and at the same time the methods of rectifier bank construction were adapted to suit this new type of element. The pace of introduction of silicon, particularly in high-voltage applications, such as traction, was limited more by the necessity to develop adequate protective systems, for example high-speed fuses, than by the unsuitability or basic unreliability of the silicon rectifier itself.[7,8]

The decade was characterized by arguments, often vehement, between equipment engineers blaming rectifier bank failures on silicon, and semiconductor engineers maintaining that uncontrolled current or voltage surges caused the trouble. By the end of the decade, such arguments had become rare, and the new techniques were fully assimilated into power engineering, giving the advantages of higher rectification efficiency and low volume, and using for the first time rectifiers with no known wear-out mechanism.

Air-cooled silicon rectifier banks in series-parallel strings were being used for motor supplies, battery charging, electroplating, electrolysis, and arc supplies for welding and furnaces.[7]

In the important field of traction, germanium rectifiers had given way to silicon for locomotive power supplies; a typical application was the use of 1·2 kV, 125 A silicon rectifiers to provide rectified power to a British Rail 4040 h.p. unit; six parallel strings of cells were used, each having four cells per arm.[8] After several years of use, it was confirmed that the silicon rectifier units in this equipment needed virtually no maintenance.

In parallel with the development of silicon power rectifiers, a development of comparable importance, that of the thyristor or controlled rectifier was taking place in the field of power control. The thyristor, depending basically on the rapid change of current gain with increasing current in p-n-p and n-p-n structures, was developed in silicon as an extension of the original p-n-p-n structure described by Moll *et al.*[9]

In the early 1960s silicon thyristors capable of handling up to about 100 A and 300 V were introduced. In the UK, thyristors suitable for use on 440 V lines appeared about the middle of the decade. These units had to withstand 1200 V peak. By 1970, the thyristor was developed to a point where a single cell could handle almost 1000 A, and withstand 2·5 kV.

The introduction of this class of device to power engineering took a course parallel to that of the silicon power rectifier, in that it took time for the device to become established in use. Over a period of about five years, the advantages of the electronic control circuitry which could be used with the thyristor became increasingly evident, and the problems of low thermal capacity and over-voltage protection were overcome as with silicon rectifiers.

Another initial difficulty in the application of the thyristor was the premature switching of the device caused by capacitive current when high rates of voltage rise were applied; this was also a defect of the early p-n-p-n switch.[9] This necessitated further development, resulting in the short-circuited emitter type of device which overcame this problem.

Turn-off problems due to minority carrier storage when using the thyristors in high frequency applications were also successfully overcome, and the frequency range of thyristors for use in inverters, for example, was greatly improved towards the end of the 1960s.

4 Small-signal Transistors

By the beginning of the 1960s a wide range of signal transistor types had been proposed and investigated. In the field of bipolar transistors germanium alloyed devices of the type described by Saby[10] had been fully developed. Variants of this design were by far the most commonly used types at that time, covering most commercial and industrial applications, and extending in frequency range up to about 5 MHz.[11,12] These transistors were also widely used in switching applications, e.g. in computers,[13] handling waveform edges with rise-times in the 1 μs region.

Such transistors were limited in frequency performance because of their basic construction, which defined the base width by alloying emitter and collector junctions from opposite sides of a wafer. Since extremely thin base width (less than 5 μm) was desirable to improve the frequency response of the device, it was necessary to control the original germanium wafer thickness to limits of about 1 μm. This proved extremely difficult to do, and already in 1960 transistor production engineers were employing methods of base width definition which did not involve the well-known hazards of subtracting one large value from another to define a small value.

The first of these was the surface barrier transistor, first described in 1953,[14] and the associated micro-alloyed device. In spite of the easier definition of small base width, and enhanced frequency response in the 50 MHz region, the poor power handling characteristics and low thermal capacity of the device were serious drawbacks.

The transistor types so far described were similar in concept from a manufacturing viewpoint in that each transistor structure was formed individually; in the case of alloy transistors, the chips were treated one at a time, as were the surface barrier chips.

A revolution in manufacturing methods was, however, under way using diffusion technology, which offered the great advantages of extremely precise dimensional control of the structure, as well as the ability to process many thousands of transistors at one time under virtually identical conditions.

Although diffusion had been used as a fabrication technique prior to 1960, notably in the construction of the successor to alloy transistors, the germanium drift transistor[15,16], and in early double-diffused transistors, it was the discovery of the oxide masking effect in silicon by Frosch and Derrick which proved to be the key to further development.[17] The effect was first used[18] to produce a silicon n-p-n transistor in which the n-type silicon chip formed the collector, and both base and emitter were diffused from one side. The emitter area was defined by oxide masking before emitter diffusion, and the collector area limited by 'mesa' etching. Devices of this type could be made with base widths less than 2 μm, allowing cut-off frequencies in excess of 100 MHz to be achieved.

In spite of the attractions of this type of construction for producing high-frequency transistors, the device suffered in switching applications from the high 'on' resistance associated with the relatively thick collector body. A solution to this problem was found by building the transistor in a thin epitaxial layer grown on to a low resistance collector body,[19,20] which produced a transistor having all the advantages of the double-diffused mesa device with good switching characteristics in addition.

Those manufacturers who, in the early 1960s, chose a diffusion-based transistor technology were to score heavily in future years over the ones who were pushing the alloy or surface barrier routes to the limits of their capabilities. The adoption of the diffusion route marked a basic change in manufacturing methods in that the formation of the transistor structure was carried out in a batch processing sequence on silicon slices rather than in individual units; the assembly was an individual operation in which emitter and base contact were thermocompression-bonded to the separate transistor chips.

The major development to confirm the superiority of the oxide masked diffusion technique for signal transistors was that of planar processing[21] in which both base and emitter areas were defined by oxide masking, and an oxide layer was used to passivate the transistor surface.

The planar method formed the basis of further small-signal transistor development during the decade. It allowed dramatic improvements to be made in both performance and cost. The definition of transistor geometries down to lateral dimensions in the order of 1 μm by planar technology resulted by 1967 in silicon transistors with a cut-off frequency in excess of 7 GHz having a power gain of greater than 4 dB at 4 GHz.[22] Similar transistors were able to switch 1 ns waveforms. By the end of the decade microwave silicon transistors with a cut-off frequency of 6 GHz and a maximum frequency of oscillation of 18 GHz were a reality. These transistors were produced using the same basic planar technology introduced in 1961, pushed to the present-day limits of geometrical resolution and processing refinement.

In large-scale production of cheap small-signal transistors for commercial applications, the main property of planar technology to be exploited in the 1960s was that of surface passivation. Transistors could be cheaply encapsulated in plastic by mass production methods, which produced devices exhibiting reliability comparable with that achievable by hermetic encapsulation.[23–25]

The planar process has shown itself to be adaptable to high-speed transistor production methods which will remove the attraction of assembly by cheap labour in the Far East. These methods have resulted in high performance plastics-encapsulated transistors being made in hundreds of millions per annum at prices as low as 3p.

Planar technology was also employed in the development of unipolar transistors of the m.o.s. type, first introduced by Hofstein and Heiman in 1963.[26] This added a high input impedance device to the armoury of the electronic equipment engineer. Although the m.o.s. device only achieved small usage compared with the bipolar planar transistor, it has formed the basis of a most important family of integrated circuits (see Section 6).

5 Power Transistors

Power transistors, usually defined as those transistors capable of operation above about 100 mA, made their appearance in about 1952, in germanium alloy form. The geometry was usually a larger version of the alloyed small-signal transistor, but with the collector soldered directly to a screwed copper heat sink allowing up to 10 W dissipation. These early devices showed rapid fall-off in current gain above 1 A[27] and were supplanted in 1957 by devices in which the indium emitter electrode was doped with gallium to improve the high-current gain.

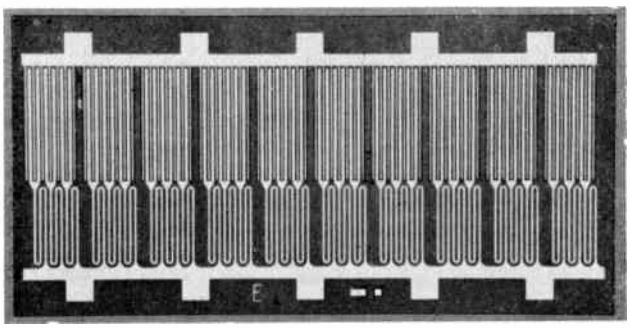
Further improvements in germanium power structures were made at the end of the 1950s with the advent of the diffused base power transistor. At about the same time, the first all-diffused silicon power transistors were appearing.[28] Emitter and base layers were diffused simultaneously over the whole surface of the device and rectifying base contacts were made by alloying through the emitter layer. These devices, although crude in construction by modern standards, showed good gain characteristics at currents up to 5 A and demonstrated the potential superiority of silicon over germanium in the power field conferred by the improved operating temperature characteristics of the silicon device.

Although the planar process was extensively used in silicon signal transistors from 1962 onwards, it was not widely adopted for silicon low-frequency power devices; most silicon power devices in commercial and industrial use are either n-p-n single diffused types or, where the highest voltage is required, based on the triple-diffused structure which made its appearance about 1965. It is possible, using silicon devices of this design, to achieve voltages in excess of 1000 V for uses such as television raster deflexion, to switch current through the line transformer.

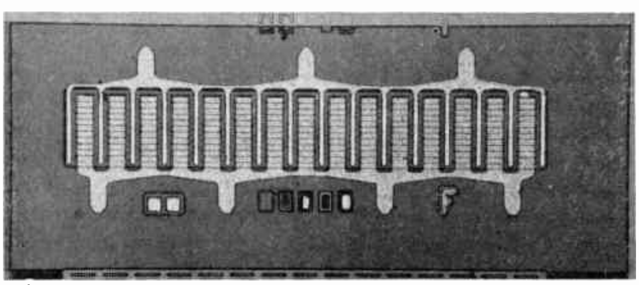
Low-frequency power transistor packaging gradually evolved during the 1960s into a fairly standardized form, and most hermetically-sealed devices are packaged in the TO3, TO36, or TO66 outlines. When the technology of silicon chip production was mastered, however, the cost of these hermetic packages was an important element in determining the device selling price and, as for the silicon small signal device, it became attractive to look for cheaper alternative systems. Plastic power packages began to appear in 1965, and have proved increasingly popular since then. This package type has exhibited a good and ever-improving reliability record, which has long been adequate for industrial applications, and which by the end of the decade was already a contender for many applications involving more demanding environmental conditions.

As with signal transistors, the silicon power transistor was pushed in performance to ever higher frequency of operation throughout the 1960s. Epitaxial processing, introduced early in the 1960s, was a major factor in reducing saturation voltage drop in power transistors, thus improving efficiency and raising cut-off frequency dramatically. At the same time, the realization that lateral voltage drop in diffused base regions restricted emitter action to the edges of the emitter structure led to the emergence of power transistor structures having more sophisticated geometries. These included the so-called snowflake and star patterns designed to provide the maximum emitter edge length in a given area.

One of the two most successful variants proved to be the inter-digitated structure in which the emitter consists of a number of connected fingers, which interlock with similar fingers forming the base contact pattern. The construction of interdigitated devices in planar epitaxial form produced a wide range of high-frequency power devices; a survey in 1968[29] illustrated the power output capabilities of this type of device, which range from 50 W at 50 mHz to 1 W at 2 GHz. Figure 3(a) shows a typical interdigitated structure capable of 78 W c.w. power at a frequency of 30 MHz.



(a) Interdigitated structure (3·56 mm × 1·78 mm) capable of 78 W at 30 MHz. The emitter ballast resistors can clearly be seen in the lower part of the structure.



(b) Overlay structure (0·5 mm × 1·3 mm) capable of 5 W at 2 GHz, employing a large number of separate emitters.

Fig. 3. R.f. power transistor geometries.

The other notably successful geometry was the 'overlay' structure introduced in 1965 by RCA in which a large number of separate emitters in a common base region is connected by metallization. The base metallization is also arranged to provide a low resistance path for base current.[30] This technique effectively parallels many small-signal devices, resulting in a power density and frequency capability characteristic of such devices, and leading to high total power at high frequency (Fig. 3(b)). From 1965 to 1970, the development of this type of device increased its power output capacity from 2 W to over 100 W at 50 MHz, and enabled 10 W to be obtained at 2 GHz.[31]

A matter of great concern during this period was the burnout of power transistors caused by the 'second breakdown' effect in which localized energy in the device leads to hot-spots. This was largely solved by the introduction of built-in series resistors (see Fig. 3(a)) in the emitter structure, which limit the local power dissipation. Present-day interdigitated and overlay structures now show good resistance to second breakdown effects, and high reliability.[32]

These developments, together with the associated special package constructions which were produced during the 1960s, have revolutionized the outlook in many equipment fields, e.g. telemetry systems, mobile communications systems for both ground and airborne use, and radar transponder systems.

6 Integrated Circuits

The 1960s have been the decade of the integrated circuit. This development in solid-state electronics will undoubtedly have a greater impact than the invention of the transistor. The transistor provided a cheaper, convenient and more reliable means of performing circuit functions than the valve, but the integrated circuit will increasingly form the foundation of future electronic system design. Without it indeed, it would not be possible to make many electronic systems of today and of future years, since it confers advantages of cost per function and system reliability without which these systems could not operate economically.

The integrated circuit concept began in 1959, when Kilby[33] demonstrated a solid-state multivibrator containing the equivalent of eight resistors, two capacitors, and two transistors. The circuit was formed on a silicon chip approximately 5 mm × 2·5 mm (0·2 × 0·1 in). The resistor areas were defined by etching and the transistors were of 'mesa' construction; some of the interconnexions were made by thermocompression-bonded gold wires. Although it was possible to construct integrated circuits in this way they were never produced in quantity; it was the advent of planar technology in 1962 which allowed such circuits to make a serious entry into the commercial field.

Naturally, this occurred first in digital computers where the repetitive form of the circuits, together with the low values of resistors and capacitors used in computer logic were ideally suited for translation into solid state form. The first circuits to appear used direct coupled transistor logic,[34] this type lending itself most easily to solid-state layout, but having the disadvantages of poor noise immunity and low fan-out. In spite of these problems, the advantages of integrated circuit design over the older discrete component designs was such that the circuits were used in a number of applications. The complexity of integration at this time was limited by manufacturing yields to a few tens of components per chip.

For a brief period, complementary emitter follower logic using p-n-p and n-p-n transistors on the same chip was examined. However, because of the very high cumulative gain and the non-saturating mode in which the circuit operates, the danger of oscillation is high, and for this reason the method was not put into large-scale production.

The first type of logic which saw widespread use was diode transistor logic, or d.t.l. This logic has a high level of noise immunity, typically 1 V, high stability, and easy loading rules. The first circuits to be introduced commercially were the Fairchild 930 range, having propagation delays in the 30 ns region. Subsequently, faster ranges were introduced, notably in the UK the Ferranti Micronor II range, with propagation delays of less than 10 ns. This type of circuit which was introduced about 1966, is still the fastest form of d.t.l. available.[35]

D.t.l. circuits are still widely used in commercial and military equipments. To obtain high operating speeds with d.t.l. it is necessary to gold-dope the circuits to kill the minority carrier storage, since the circuit operates by driving the transistors into saturation. In order to maintain transistor gains at these low minority carrier lifetimes the transistors must be built with very thin base regions, and emitter-collector short-circuits were often a limiting factor in determining manufacturing yield. Nevertheless, by 1967 the number of components per chip was approaching 100.

From the point of view of the manufacturer, the decade was a period of struggling to improve yields to an economic level at a given degree of complexity, and when this was achieved, the struggle began all over again at the next higher level of complexity. There is, as yet, no sign of an end to this progression!

One of the most successful variants of d.t.l. has been transistor-transistor logic or t.t.l., which provides a low impedance output to the supply in both on and off states. The gate inputs are via multi-emitters of a transistor; the low impedance fast waveform edge transmitted to the output stage removes stored charge in the output transistor and enhances the circuit operating speed.

This type of circuit has had enormous success, and was made in the well-known 7400 range by at least twenty manufacturers during the 1960s. The complexity increased towards the end of the decade, and more and more functions, e.g. full adders, shift registers, decade counters, etc., were added as the yield problems were solved and several hundred components per chip became a reality.

The enormous popularity of t.t.l. which was mainly sold in cheap plastic moulded packages, unfortunately rebounded on the integrated circuit manufacturers in 1969 and 1970. At this time there was a cut-back in military and space technology in the U.S.A. which coincided with a drop in demand for the circuits in commercial applications such as the computer industry. Many manufacturers, used to an expansion rate of at least 25% per annum, found themselves with huge t.t.l. stocks, and liquidity problems forced them to dispose of these stocks, both in the USA and overseas, at prices which in many cases were less than production costs. The selling price of some circuits dropped by a factor of 10 within twelve months. Inevitably, many manufacturers who depended on t.t.l. for most of their output were forced out of the business, and thousands of engineers and scientists, particularly on the west coast of America, found themselves without jobs.

The successor to t.t.l. for high-speed computer work, which first saw commercial use about 1967, is emitter coupled logic in its various forms.[36] These circuits have demonstrated the fastest switching speeds so far available (1 ns and less) and in order satisfactorily to incorporate them into central processors it has been necessary to adopt transmission line interconnexion techniques.

All the above circuit types depend on bipolar technology, which was limited in the complexity of circuitry it could handle because of yield problems. New forms of bipolar circuit are now emerging which show the promise of very high manufacturing yield, and will allow large and complex sub-systems to be made on one chip.

One of the most promising of these is collector diffusion isolation (c.d.i.), which first appeared at Bell Laboratories in 1969.[37] Although in its original form the technique was limited, particularly in operating voltage rating, it was subsequently developed in the UK[38] to overcome these limitations, and allow high speed (5 ns) systems to be constructed which have in excess of 7000 components on a chip less than 0·25 in square (see Fig. 4).

Thus, in the 1960s, bipolar digital circuits progressed from a chip containing a single logic gate to a chip on which a complete small computer central processor can be made.

Bipolar linear integrated circuits developed later than their digital counterparts, and indeed the linear field showed little activity before 1965. In attempting to develop early linear circuits, the designers were limited by the component values which could be produced in integrated circuit form. Resistors, and particularly capacitors, were limited in value and the temperature coefficients of these components were considerably higher than those of their discrete equivalents. In addition, resistor tolerances were limited to about ±20%.

It was some time before circuit engineers learnt to adopt a new philosophy of linear circuit design which used the advantages of the new medium; these include closer tracking of component parameters when they are formed close together on a chip, and the extreme cheapness of active devices in integrated circuit form. When the new methods of designing linear circuits in solid state form were adopted, the circuits so produced were often more complex than their discrete component equivalents, but excellent performance characteristics could be achieved.[39,40] A notable example was the operational amplifier in its various forms, a type of circuit which has seen widespread use and is by far the most popular linear integrated circuit to have been produced.

The μA709 operational amplifier contains 14 transistors and 15 resistors, about three times the number of components in the discrete version. Its operational characteristics, however, are at least equivalent to the discrete assembly at very much lower cost. This circuit has been widely used in analogue computing and communications applications; in computing it can add, subtract, differentiate and integrate. In communications it can be designed to give a range of response curves, including that of the wide-band video amplifier.

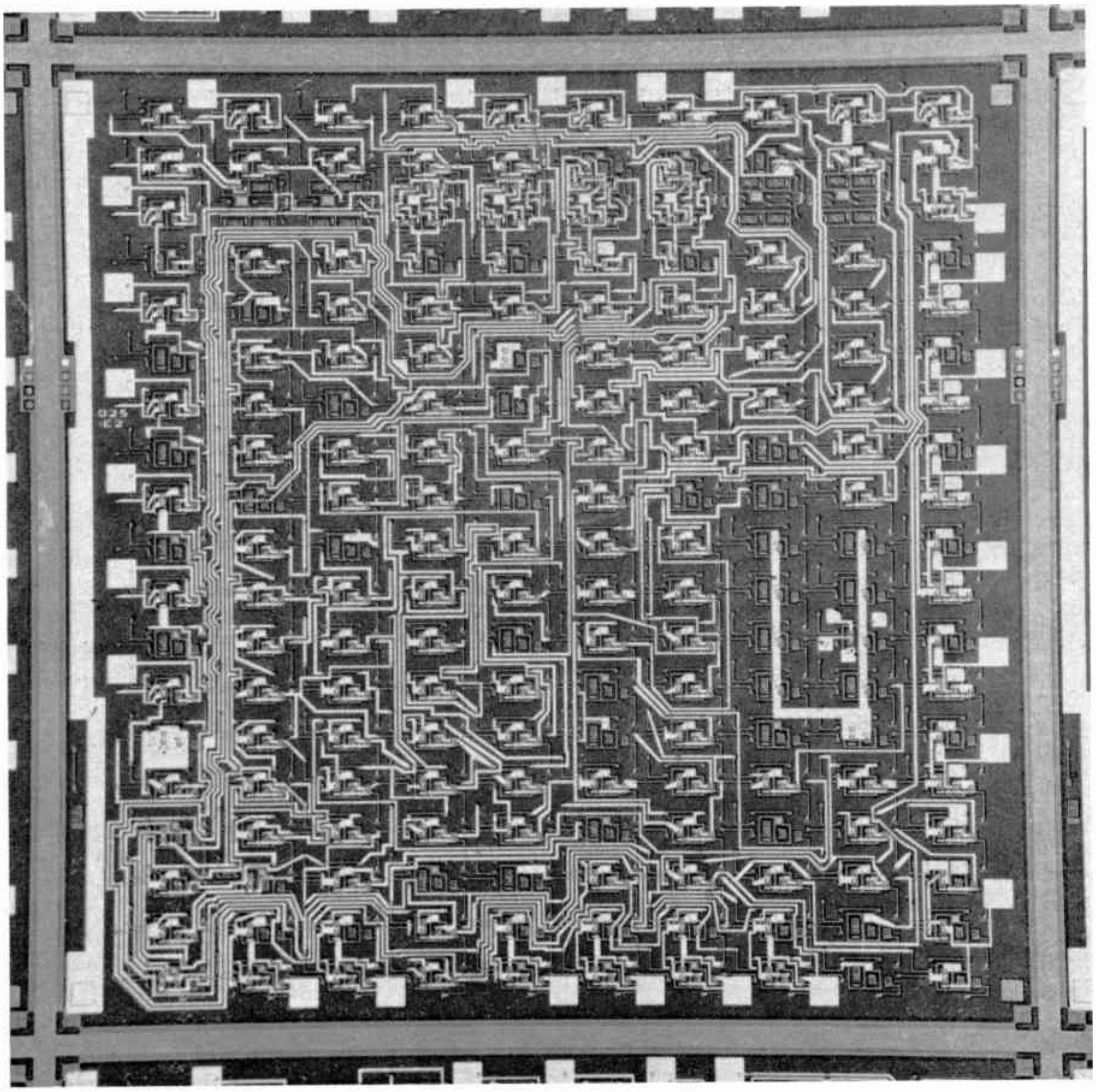


Fig. 4. Typical high density c.d.i. integrated circuit. The example shown is an uncommitted logic array containing 200 gates, which can be interconnected to satisfy any one of a large number of users' requirements. (Chip size 3·5 mm × 3·5 mm).

In more recent years the use of linear design techniques has been extended to cover both r.f. and a.f. amplification, and many high-quality a.f. amplifiers are now available in power packages handling up to 10 W of output power, together with integrated low-noise pre-amplifiers for use in high-fidelity sound reproduction units.

At the end of the decade the impact of linear integrated circuits began to be felt in the field of television signal processing. Circuits which became available at that time in low-cost plastics-encapsulations include integrated television sound circuitry, from video pre-amplifier to loudspeaker, contained in a single package.

From 1965, the use of the unipolar or field effect device[26] in integrated circuits increased greatly. By far the most commonly used type of device has been the metal-oxide-silicon or m.o.s. device, in which conduction between source and drain regions in the silicon body is modulated by a metal gate insulated from the underlying region between source and drain by a thin oxide film.

The m.o.s. device, which has a very high input impedance, is simpler in construction than the bipolar transistor, usually occupies less area, and when used in an integrated circuit is self-isolating from neighbouring devices. This allows a device packing density considerably higher than in conventional bipolar integrated circuits (see Fig. 5), although the new c.d.i. system described previously can rival m.o.s. in this respect.

The conducting channel between source and drain may be either p or n-type; from the viewpoint of operating speed and size reduction, the n-channel is to be preferred. The n-channel process is, however, more difficult to control, so that it was not much used initially. More recently, the c.m.o.s. process, using both p and n-channel devices has been introduced; the circuit is arranged so that the sources of the n-channel devices are held at earth

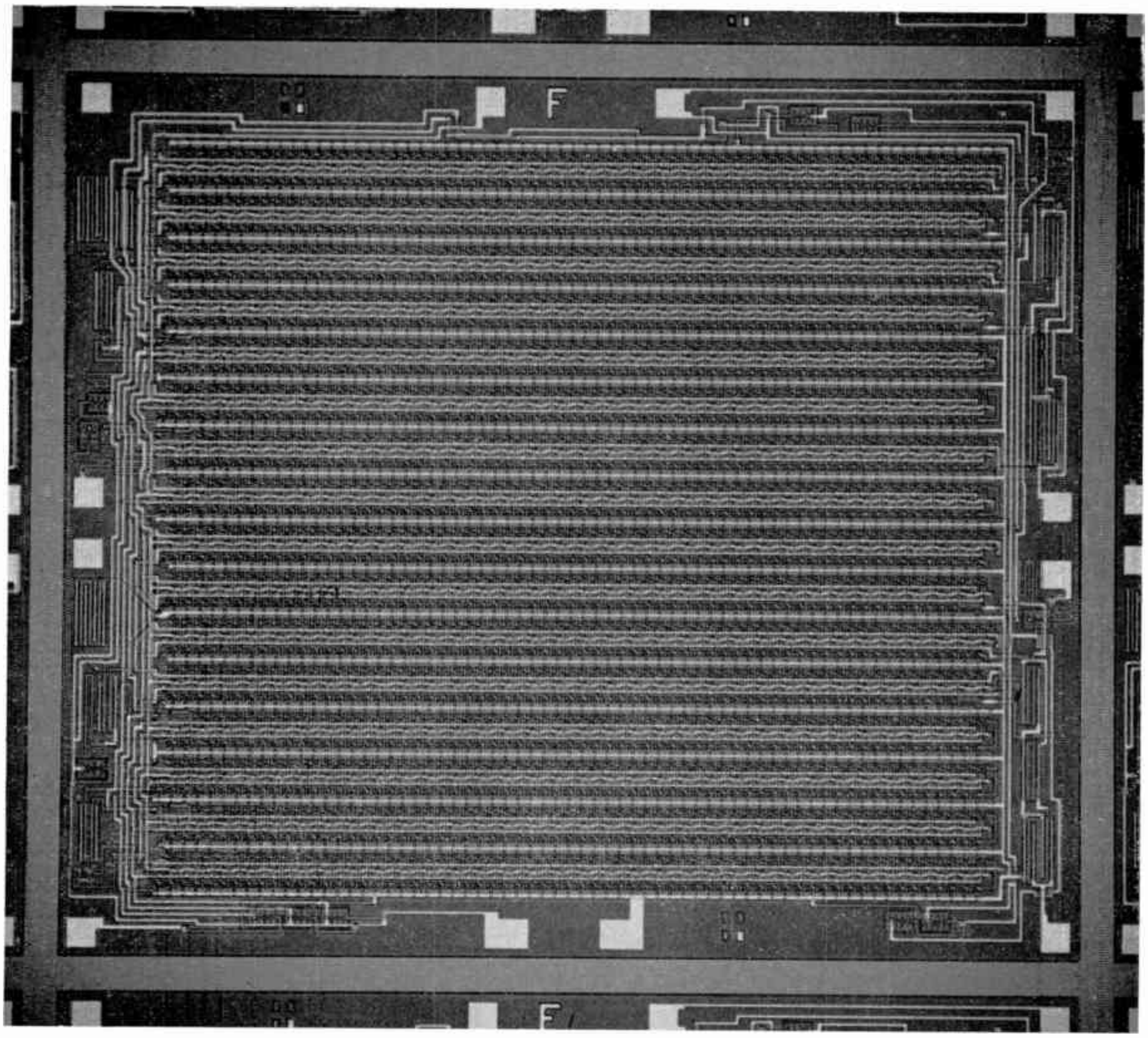


Fig. 5. Typical complex m.o.s. circuit. The example shown is a silicon gate shift register containing 1024 bits. (Chip size 3·5 mm × 3·5 mm).

potential, and this simplifies the stability problems. Almost all m.o.s. circuitry made in the 1960s, even the predominating p-channel devices, exhibited surface stability problems to some extent. This was basically due to the fact that the surface concentrations of impurities in these devices are substantially lower than those in bipolar circuits and unwanted surface charges can thus more easily affect the operation of the device. Better surface characteristics can be obtained by using a silicon nitride insulating layer in combination with silicon oxide. This is usually referred to as the m.n.o.s. transistor and is characterized by a lower threshold voltage. In addition, the m.o.s. device is easily damaged by accidental build-up of static charge which can take place during handling, and precautions against this during assembly of the devices have proved necessary.

In spite of these disadvantages, the usefulness of the m.o.s. circuit in applications requiring large repetitive arrays has led to its widespread use. Computer storage has been a particularly suitable application, and random access memories using the m.o.s. circuit were becoming increasingly common towards the end of the 1960s.

The type 1103 random access memory with 1024 bits has become an industry standard, and allows relatively high speed storage to be obtained at a price to the equipment maker of about 0·2p per bit. M.o.s. and m.n.o.s. technology will undoubtedly make increasing inroads into computer storage systems as the cost is further reduced.[41]

By 1969, an exciting new development in the use of m.o.s. and m.n.o.s. circuits in telephone systems was also beginning to come into commercial use in the UK. This involves the use of four-phase logic in the design of a wide range of circuits for subscriber and exchange equipments, which promises enormously to expand the role of the telephone, making possible for example storage of frequently used numbers in the subscriber's instrument, and the use of credit cards via telephone lines.

7 References

1. Pearson, G. L. and Sawyer, B., 'Silicon p-n junction alloy diodes', *Proc. IRE*, **40**, pp. 1348–51, August 1952.
2. Prince, M. B., 'Diffused p-n junction silicon rectifiers', *Bell Syst. J.*, **35**, pp. 661–84, December 1955.

3. Olson, K. H. and Robillard, T. R., 'New miniature glass diodes', *Bell Laboratories Record*, **45**, pp. 13–15, January 1967.
4. Roualt, C. L. and Hall, G. N., 'A high voltage, medium-power rectifier', *Proc. IRE*, **40**, pp. 1519–21, August 1952.
5. Klein, N. 'Thermal stresses and fatigue in silicon power rectifiers', *IEEE Trans. on Communications-Electronics*, No. 71, pp. 208–19, March 1964.
6. Shields, J., 'An advance in power rectifiers', *New Scientist*, No. 391, pp. 434–36, May 1964.
7. Bennell, F. T., 'D.c. supplies met by silicon rectifiers', *Electrical Times*, **154**, No. 4003, pp. 343–6, September 1968.
8. Sommerschield, J. G., 'Application of rectifiers to British Railways a.c. motive power', *Proc. IEE*, **116**, No. 2, pp. 247–60, February 1969.
9. Moll, J. L., Tanenbaum, M., Goldey, J. M. and Holonyak, N., 'p-n-p-n transistor switches', *Proc. IRE*, **44**, pp. 1174–82, June 1956.
10. Saby, J. S., 'Fused impurity p-n-p junction transistors', *Proc. IRE*, **40**, pp. 1358–60, July 1952.
11. Herold, E. W., 'New advances in the junction transistor', *Brit. J. Appl. Phys.*, **5**, pp. 115–26, April 1954.
12. Ebers, J. J., 'Alloyed-junction transistor development', *Bell Laboratories Record*, No. 1, pp. 8–12, January 1956.
13. Henle, R. A. and Walsh, J. L., 'The application of transistors to computers', *Proc. IRE*, **46**, pp. 1240–54, June 1958.
14. Bradley, W. E., 'The surface barrier transistor' (series of 5 papers), *Proc. IRE*, **41**, pp. 1702–20, October 1953.
15. Kroemer, H., 'Transistors 1', *RCA Rev.*, **17**, p. 202, 1956.
16. te Winkel, J., 'Drift transistor—simplified electrical characterisation', *Electronic and Radio Engineer*, **36**, pp. 280–8, August 1959.
17. Frosch, C. J. and Derrick, L., 'Surface protection and selective masking during diffusion in silicon', *J. Electrochem. Soc.*, **104**, p. 547, 1957.
18. Aschner, J. F., Bittman, C. A., Hare, W. F. J. and Kleimack, J. J., 'A double diffused silicon high frequency transistor produced by oxide masking techniques', *J. Electrochem. Soc.* **106**, pp. 145–7, May 1959.
19. Theuerer, H. C., Kleimack, J. J., Loar, H. H. and Christensen, H., 'Epitaxial diffused transistors', *Proc. IRE*, **48**, p. 1642, September 1960.
20. Valdes, L. B., 'Characteristics of silicon epitaxial transistors', *Solid State J.*, **2**, p. 33, November 1961.
21. Hoerni, J. A., 'Method of manufacturing semiconductor devices', Institute of Radio Engineers' Electron Devices Meeting, Washington D.C., October 1961.
22. 'Reducing internal dimensions improves silicon transistor performance', *Bell Laboratories Record*, **45**, pp. 13–15, January 1967.
23. Sideris, G., 'Belting out plastic transistors on mechanized assembly lines', *Electronics*, **39**, No. 2, pp. 84–90, January 1966.
24. 'The Epoxy Encapsulated Transistor'. Ferranti Publication (Electronic Components Division), March 1967.
25. Latham, G. R., 'Silicon planar reliability and the future, Parts 1 and 2', *Electronic Components*, **7**, May & June 1966.
26. Hofstein, S. R. and Heiman, F. P., 'The silicon insulated gate field-effect transistor', *Proc. IEEE*, **51**, p. 1190, September 1963.
27. Roka, E. G., Buck, R. E. and Reiland, G. W., 'Developmental germanium power transistors', *Proc. IRE*, **42**, pp. 1247–50, August 1954.
28. Aldrich, R. W., Lanzl, R. H., Maxwell, D. E., Percival, J. O. and Waldner M., 'An 85-watt dissipation silicon power transistor', *IRE Trans. on Electron Devices*, **ED-5**, pp. 211–5, October 1958.
29. Tatum, J. G., 'Fingers in the die', *Electronics*, **41**, pp. 93–7, February 1968, (Special Report).
30. Carley, D. R., McGeough, P. L. and O'Brien, J. F., 'The overlay transistor, part 1: new geometry boosts power', *Electronics*, **38**, pp. 71–7, August 1965.
31. Carley, D. R., 'A worthy challenger for r-f honors', *Electronics* **41**, pp. 98–102, February 1968, (Special Report).
32. Johnson, J. H. and Mallinger, M. J., 'You can depend on todays' r-f power transistors', *Electronics*, **44**, pp. 90–3, September 1971.
33. Kilby, J. S., 'Semiconductor solid circuits', *Electronics*, **32**, pp. 110–1, August 1959.
34. Harris, J. R., 'Direct-coupled transistor logic circuitry'. *IRE Trans. on Electronic Computers*, **EC-7**, pp. 2–5, March 1958.
35. Bagnall, P. J., 'The Micronor II exercise (parts I & II)', *Electronic Components*, **7**, December 1966 and **8**, January 1967,
36. Bilous, O., Feinberg, I. and Langdon, J. L., 'Design of monolithic circuit chips', *IBM J. Res. Developm.*, **10**, September 1966.
37. Murphy, B. I., Glinski, V. J., Gary, P. A. and Pedersen, R. A., 'Collector diffusion isolated integrated circuits', *Proc. IEEE*, **57**, pp. 1523–7, September 1969.
38. Grundy, D. L., Bruchez, J. and Down, B., 'Collector diffusion isolation packs many functions on a chip', *Electronics*, **45**, pp. 96–104, July 1972.
39. Camenzind, H. R. and Grebene, A. B., 'An outline of design techniques for linear integrated circuits', *IEEE J. Solid-State Circuits*, **SC-4**, No. 3, pp. 110–21, June 1969.
40. Widlar, R. J., 'Some circuit design techniques for linear integrated circuits'. *IEEE Trans. on Circuit Theory*, **CT-12**, No. 4, pp. 586–90, December 1965.
41. Mazda, F. F., 'The components of computers, part 4: large scale integration—the r.a.m.', *Electronic Components*, **13**, No. 18, pp. 876 *et seq.*, September 1972.

Manuscript received by the Institution on 2nd November 1972. (*Paper No. 1491/CC 155.*)

UDC 621.382

The technology of semiconductor materials preparation

J. C. BRICE, M.A., Ph.D.*

and

B. A. JOYCE, B.Sc.*

SUMMARY

This paper describes the preparation of semiconductor single crystal material in bulk and thin films. Particular emphasis is placed on silicon and gallium arsenide, and the growth methods are described in relation to the electrical characteristics which are required for device fabrication.

*Mullard Research Laboratories, Redhill, Surrey.

1 Introduction

The development of the transistor over the last twenty-five years has been accompanied by great advances in the technology of semiconductor materials preparation. The technical and economic pressures created by the new industry have resulted in new processes being invented and old ones being developed to levels which were inconceivable in 1948. This paper looks at some of the processes used to prepare materials. It does not aim to give a complete history of the subject but rather concentrates on those aspects which are currently important or which are likely to be important in the future.

The great majority of present-day transistors are made from silicon although substantial numbers of germanium devices are still made. In the future, it seems likely that f.e.t.s based on gallium arsenide may also be produced. Accordingly, the paper concentrates on these three materials.† In order to obtain high-performance devices with reasonable yields, the materials used must meet stringent specifications. These specifications concern the structural and electrical properties, which are completely interdependent. The essential criterion is that the material shall be in the form of a single crystal containing the minimum number of crystallographic defects, and having a precisely controlled concentration of ionized impurities over the approximate range 10^{13}–10^{20} atoms cm^{-3}. The electrical specification will define the required majority carrier concentration, distribution and mobility, and the minority carrier lifetime; all of these parameters are influenced by structural factors, and it is essential to minimize crystallographic imperfections if the electrical properties are to be sufficiently well controlled. Modern technology has reached the stage where dislocations, stacking faults and micro-twins can be entirely eliminated from the 'as-grown' crystals, but other types of structural imperfections, referred to generally as point defects, and including principally lattice vacancies and extraneous impurity atoms, still cause problems in device fabrication.

In general terms, imperfections in the crystal lattice can affect four parameters crucial to device fabrication and behaviour.

(i) Majority carrier properties can be influenced as the result of increased scattering and mobility reduction.

(ii) Minority carrier lifetime can be reduced by the introduction of trapping and recombination centres.

(iii) Impurity diffusion rates can be enhanced by lattice defects.

(iv) Carrier concentrations may be changed as the result of impurity atom-defect interactions.

The remainder of this paper will consider methods for the preparation of high-perfection single crystal semi-

† Silicon, germanium and gallium arsenide are also the materials from which the majority of diodes are made. Some other Group III–V compounds and mixed crystals of these materials have uses in light-emitting diodes. The other materials (zinc sulphide, selenide and oxide, cadmium sulphide and indium antimonide) are used in significant amounts only for photo-detectors. Overall, about 83% of semiconductor devices are made from silicon, 16% from germanium and 1% from the other materials.

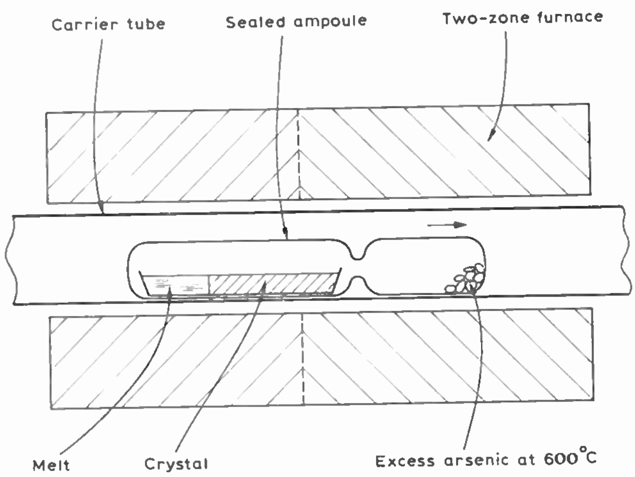


Fig. 1. The growth of gallium arsenide by the horizontal Bridgman method. The carrier tube is moved relative to the furnace at a rate of about 5 mm/h.

conductor material in bulk and thin film form, which enable these electrical effects to be controlled.

2 The Growth of Bulk Single Crystals

The starting point for the manufacture of nearly all transistors is a block of material cut from a bulk semiconductor crystal. Bulk crystals can be grown in many ways (see, for example, the reviews by Laudise,[1] Harper,[2] Gilman[3] and Brice[4]), but, in practice, bulk semiconductor crystals are almost invariably grown from their melts. This class of methods can produce large crystals of reasonable quality more rapidly than any of the other methods and so is the most economic. In some cases, e.g. alloy devices and m.o.s.t.s, the bulk crystal is used directly, in others it is more convenient to deposit one or more layers of single crystal onto the bulk crystal and make the active parts of the device in these layers. This section describes the processes for the production of bulk crystals. The methods for thin layer production are discussed in the next section.

Each of the materials has its own characteristics and these have considerable influence on the methods selected for various purposes. Germanium melts at 937°C and can be melted without significant contamination in either silica or carbon containers provided that a suitably inert atmosphere (e.g. hydrogen, helium, argon, nitrogen or vacuum) is maintained over the melt. Silicon melts at ~1410°C and is much more reactive: all the known readily-available container materials react to some extent; silica appears to be the best choice but some reaction occurs and oxygen can always be detected in crystals grown from silica vessels. Silicon also requires an inert atmosphere. Technologically, gallium arsenide is the most difficult material. It melts at 1240°C but, to prevent decomposition, the atmosphere over the melt must contain arsenic at a fairly high pressure—the solid and liquid with the same composition are in equilibrium with an atmosphere having an arsenic partial pressure of 720 torr. The choice of a material to contain the melt is difficult: all the readily-available ones cause some contamination;[5] silica is a common choice but the amount of silicon entering the crystal depends strongly on the arsenic pressure over the melt—higher pressures result in less contamination. The nitrides of boron and aluminium are also useful.

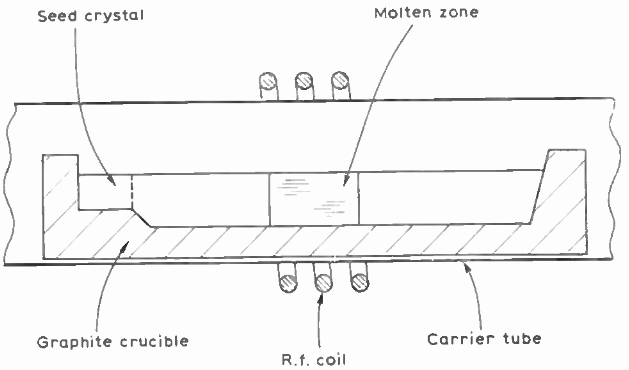


Fig. 2. The production of a germanium crystal by zone levelling.

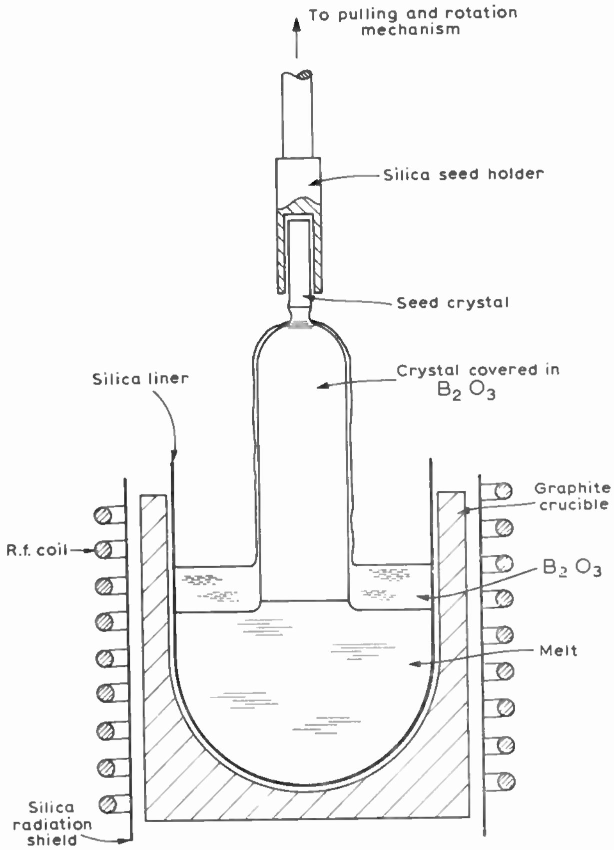


Fig. 3. Encapsulated Czochralski growth of gallium arsenide. The system is enclosed in a steel pressure vessel. Growth is initiated by dipping the seed crystal into the melt. After a pause, the seed is raised at say, 3 cm/h and rotated at 25 rev/min. For silicon or germanium the method would be the same but the boric oxide layer would be omitted.

Methods for growth from the melt can be grouped into three classes. In the first both the melt and the crystal are in contact with a crucible (see Figs. 1 and 2). In the second, the melt but not the crystal touch the container (see Figs. 3 and 4), and in the third group, no container is used—the melt is supported on its own solid by surface tension (see Figs. 5 and 6). These methods all have their own particular advantages and disadvantages.

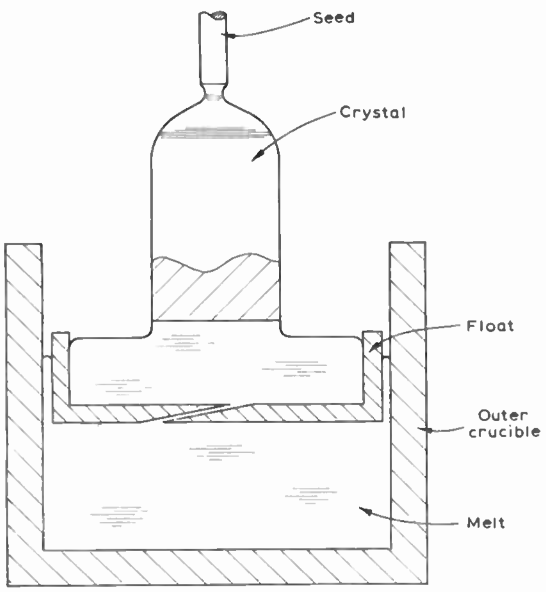


Fig. 4. Floating crucible growth of germanium. The method is not applicable to silicon or gallium arsenide because the crucible and the float can only reasonably be made from graphite. Heat can be supplied either by induction heating or by radiation from a resistively heated element made from graphite or molybdenum.

The methods shown in Figs. 1 and 2 for example have the economic advantage that they require little operator attention but they have the disadvantage that because the crystal cools in contact with the crucible, stress due to differential contraction may reduce the crystallographic quality. The process shown in Fig. 1 is a development of the technique first used by Bridgman[6] and has been widely used for gallium arsenide. The most obvious disadvantage of the method is that as the crystal grows, the volume of the melt decreases and this makes the growth of uniformly doped crystals almost impossible. If a dopant is present in the original melt at a concentration C_0, then the concentration in the initial portion of the crystal is KC_0 where K is the segregation coefficient which is seldom equal to unity. As a result, the concentration in the melt changes as the crystal grows and it can be shown that when a fraction, g, of the melt has been solidified the concentration in the crystal is

$$C_c = KC_0[1-g]^{K-1} \qquad (1)$$

since with the exception of boron in silicon ($K = 0{\cdot}8$) and boron in germanium ($K = 17$), the values of K in semiconductor systems tend to be small ($\lesssim 0{\cdot}1$), C_c changes rather rapidly with g. Burton *et al.*[7] showed that as the amount of stirring is decreased and the growth rate increased, the effective value of K tends to unity. Unfortunately, these conditions produce other problems: the growth face becomes unstable because of constitutional supercooling[8] and second phase inclusions are formed. This problem, is also present in the method shown in Fig. 3.

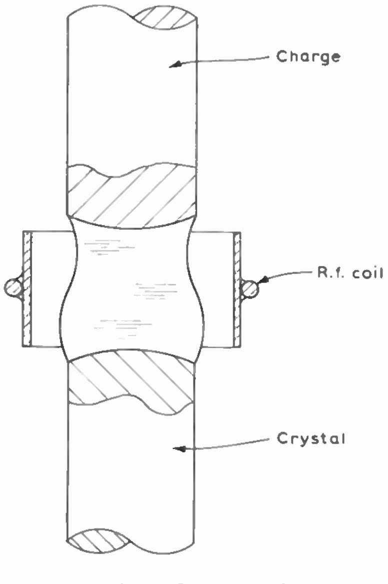


Fig. 5. Floating zone refining. If the crystal and charge have the same diameter, they are both moved at the same rate. Differences in diameter require different rates. Stirring can be introduced by rotating either the crystal or the charge or both.

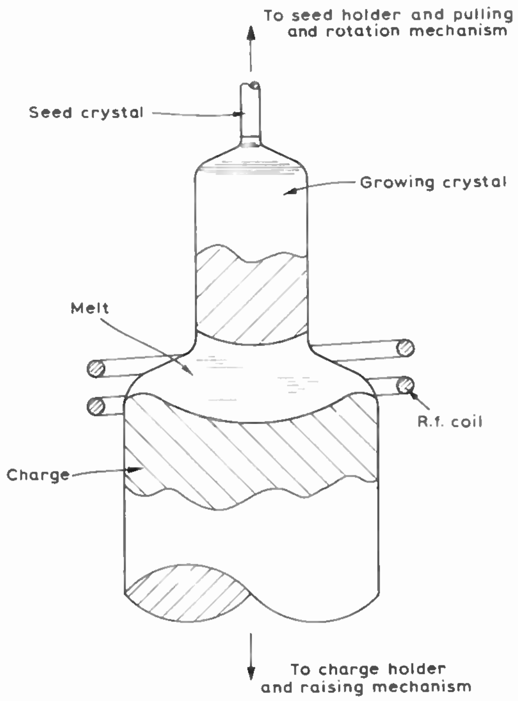


Fig. 6. Pedestal pulling of silicon.

The methods shown in Figs 2, 4, 5 and 6 have essentially constant volumes for their melts, which are replenished either from a reservoir of liquid (Fig. 4) or by melting a solid charge. If the liquid has an initial concentration C_0, the melt volume is V and the replenishing material contains dopant at a concentration C_R, then

$$C_c = C_R-[C_R-KC_0]\exp(-Kx) \qquad (2)$$

where x is the ratio of the volume of the crystal to the volume of the melt. Even if $C_R = 0$, it can be seen that for small values of K, substantial amounts of crystal can be grown with only small changes in C_c. In the ideal state C_R is adjusted to equal the desired value of C_c and C_0 is made equal to C_R/K.

One common difficulty with obtaining a uniform impurity distribution in melt-grown crystals of semiconductors is that different faces grow by different mechanisms[9] and this affects the value of the distribution coefficient:[10] in particular, the effective value on (111) faces is usually larger than on other faces so that, if the growing interface includes a (111) face, a region with different impurity concentration can be found. The problems connected with impurity distributions have been reviewed elsewhere.[11]

In order to avoid the problems caused by crucible-crystal contact, the method shown in Figs. 3 and 4 is often used. This is a development of the technique due to Czochralski.[12] Until recently, most workers used continuous visual monitoring of the process, which makes it expensive. A number of workers[13–17] have, therefore studied the automation of the process and have achieved considerable success with systems which either look at the shape of the liquid surface (which changes with the radius of the crystal) or which use television systems to look directly at the radius of the crystal. The general aim is to maintain a constant radius so that the control systems compare the radius at any time with a pre-set standard and produce an error signal which effects a correction by changing either the pulling speed, rotation rate or melt temperature. (An increase in any of these parameters decreases the radius of the crystal.) A good control system not only decreases the number of man-hours involved in growing a crystal, but also enables more perfect crystals to be grown. A human operator cannot distinguish a small change in diameter and therefore makes larger corrections than a machine. Large changes in the growth conditions produce inhomogeneities in the crystal and may introduce other defects, such as regions with enhanced vacancy concentrations or even dislocations.

Dislocations in crystals can originate from many sources which are discussed in Reference 4. In pulled crystals, the most important effects are propagation of dislocations from the seed crystal and dislocations caused by thermal stresses in the crystal. The growth of a long thin neck at the start of growth will usually eliminate dislocations which originate in the seed crystal or at the place where the crystal is joined to the seed. Dislocations resulting from thermal stress can be avoided by selecting the growth conditions to keep the growth interface fairly flat and by restricting the heat losses from the crystal, for example, by the use of a radiation shield.[18–19] Note that eliminating dislocations can create problems with strain and vacancy clusters which do not often occur in dislocated crystals.

One problem with pulling silicon crystals is the contamination of the crystals with oxygen which results from the attack of the crucible by the melt. The amount of oxygen can be decreased by reducing the rate of rotation of the crystal but some rotation is essential. To eliminate the problem of crucible contamination the methods shown in Figs. 5 and 6 were developed. Figure 5 shows floating zone refining[20, 21] and Fig. 6 shows the derived method of pedestal pulling. With both methods, the melt is in contact only with its own solid so that there are no problems of contamination. The main limitation is that the length of the liquid zone is limited by the radii of the two solid rods and the ratio of surface tension to density. This ratio is much larger for silicon than for most other materials, so that these techniques are almost always used for silicon. In order to produce dislocation-free crystals, it is again necessary to grow a thin neck and to control the heat flow. The control of heat flow is difficult and much ingenuity is needed (see, for example, Chapter 8 of Reference 4). If dopants are required they may be placed in the molten zone initially or added to the gas atmosphere in the form of a volatile compound such as phosphine.

Over the years a large volume of empirical data has been accumulated and some convincing theoretical descriptions of the various processes involved have been formulated. From this it is possible to predict the likely result of changing a parameter of the growth process. Table 1 gives some examples.

Table 1

The effects of changing the parameters of a growth process

Parameter	Effect of increasing the parameter
Temperature gradient	Increases interface stability. Increases dislocation density. Causes inhomogeneity above a critical limit.
Growth rate	Decreases interface stability. Increases dislocation density. Increases vacancy concentration. Decreases cost.
Lateral dimensions of the crystal	Increases dislocation density. Increases vacancy concentration. Decreases cost.
Accuracy of temperature control	Increases homogeneity. May decrease dislocation density.
Stirring	Increases interface stability. Increases homogeneity up to a limit above which turbulence causes inhomogeneity.

This section has given a very brief description of the methods used to produce bulk semiconductor crystals. For a much fuller review see Reference 4. Typical products vary with the material but it is not unreasonable to expect the best processes to give:

(i) germanium and silicon crystals 12 cm in diameter;

(ii) dislocation free silicon crystals with low oxygen content 6 to 8 cm in diameter with resistivity variations of less than 10–20%;

(iii) gallium arsenide crystals 3 cm in diameter which are dislocation free.

3 The Production of Thin Layers

Thin single crystal layers of the various materials can be grown either from the vapour or from solution. Both techniques are useful. The perfection of the layers produced is always related to the perfection of the substrates. It is important therefore to prepare substrates carefully and to ensure that the prepared surfaces are clean and free from damage. It is nearly always beneficial to arrange that some of the substrate is dissolved away before growth commences.

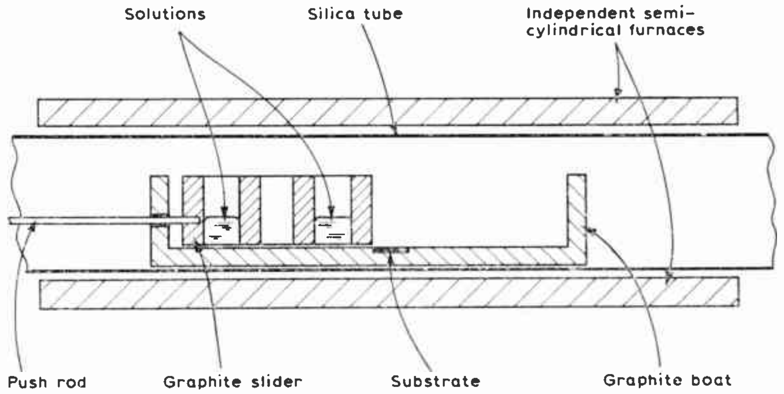


Fig. 7. A slider system for the growth of thin layers from solution. By means of the push rod the first solution can be moved over the substrate. After a period of cooling to obtain growth, this solution can be replaced by the second one and a different layer grown. Note that in order to produce a temperature gradient in the solution the lower furnace is kept about 50 degC cooler than the upper one.

3.1 *Solution Growth*

Crystals of germanium have been grown from solution in indium or tin[22] but in practice solution growth is only of importance for gallium arsenide.[23,24] Several systems have been used. The most popular are the processes in which the substrate is lowered into a crucible containing the solution and then removed when growth is completed. Alternatively the solution and the substrate can be placed in a long boat which is tilted to bring the solution into contact with the substrate for the growth period. A third technique is shown in Fig. 7. This method allows successive layers of different materials to be grown. In all cases the growth occurs because the solutions are cooled. The rate of growth depends on the rate of cooling, the area of the substrate and the volume of the solution. Generally, the slowest rates of cooling produce the least contamination. Typically an undoped gallium arsenide layer produced by cooling a solution in gallium at 0·25 degC per minute for 30 minutes might be 20 μm thick and have a carrier concentration of about 10^{15} cm^{-3} and a room temperature mobility of about 7000 cm^2 V^{-1} s^{-1}. Adding 6% of tin to the gallium can produce layers with 10^{18} carriers cm^{-3}. The principal limitation of the method is that the thickness of the layer varies over the slice. It is possible to reduce this variation by adjusting the temperature distribution but this can be tedious.

3.2 *Vapour Phase Growth*

Vapour phase growth techniques are used in semiconductor technology for the preparation of thin films of metals, insulators and semiconductors, but the most important process is the growth of single crystal semiconductor films on single crystal substrates of the same material which will be the main topic discussed in this section. The subject has been reviewed by Newman,[25] Joyce[26] and Francombe and Johnson.[27] It is a particular case of the phenomenon of epitaxy, or the growth of a film on a crystalline substrate such that a crystallographic relationship exists between them. In general, film and substrate will be of different materials, but in semiconductor technology they are almost always the same (Si on Si, GaAs on GaAs,) and the phenomenon is then more correctly referred to as autoepitaxy or homoepitaxy. In practice, however, the term epitaxy is also extensively used for this case.

The importance of epitaxial growth in semiconductor technology is directly related to the comparative ease with which impurity concentrations in the films can be controlled. This leads to many applications: p-n junctions can be produced between layer and substrate, films can be produced having much lower impurity concentrations than can be achieved in bulk crystals, which is of particular importance in GaAs technology, but the most important application is the growth of films having relatively low impurity concentrations on more highly-doped substrates of the same conductivity type. In this way the series resistance associated with the substrate can be reduced without any other change to the device characteristics, and it is one of the basic steps in the production of bipolar transistors and integrated circuits.

Two basic methods are available for the preparation of epitaxial semiconductor films: either the atoms can be supplied directly to the substrate surface by evaporation in vacuum from a source of the same material, or they can be supplied as volatile compounds which undergo chemical reaction at the heated substrate surface to yield the film material as one reaction product, all other products being volatile. Because of its much wider technological importance, this latter method, known as chemical vapour deposition, will be considered in greater detail. Review articles have been written by Schwartz,[28] Feist *et al.*,[29] Li[30] and Joyce.[31]

3.3 *Chemical Vapour Deposition*

The reasons for the importance of chemical vapour deposition to the technology of semiconductor materials may be summarized as follows:

(i) Growth can be obtained at temperatures much lower than the melting point so that contamination problems are reduced, and impurity diffusion rates within the growing film minimized.

(ii) Excellent control of layer thickness can be achieved.

(iii) Doping elements can be added in a precisely controlled manner over a very wide concentration range ($\sim 10^{13}$–10^{20} atoms cm^{-3}).

(iv) Material of high crystallographic perfection can be grown simply, without recourse to ultra-high-vacuum techniques.

Three basic reaction systems have been used to prepare semiconductor films. These are (i) pyrolysis of hydrides, (ii) hydrogen reduction of halides and (iii) disproportionation reactions. For some compound semiconductor

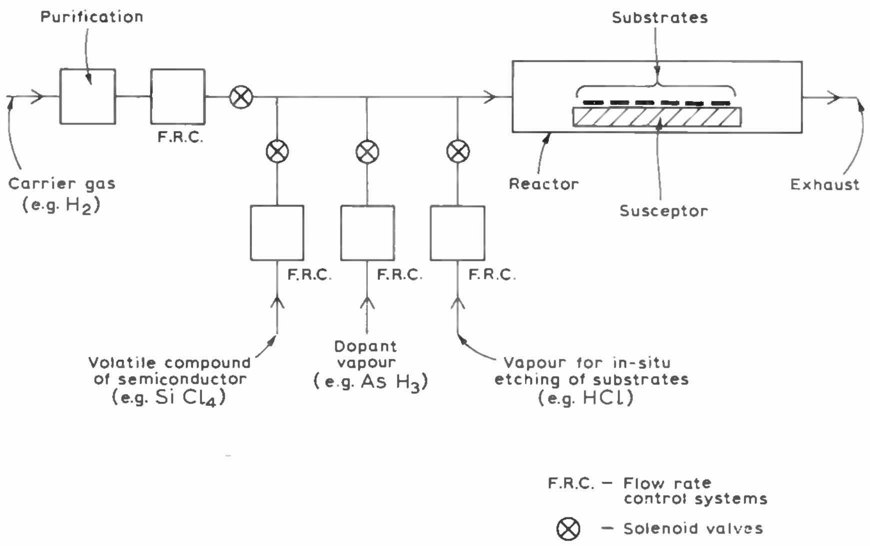


Fig. 8. Schematic illustration of the basic elements of a chemical vapour deposition system.

films, e.g. GaAs, methods (i) and (ii) have been used in conjunction. Typical reaction systems are represented schematically by the overall equations shown below.

(1) (pyrolysis)

$$SiH_{4(v)} \rightarrow Si_{(s)} + 2H_{2(v)}$$

(2) (reduction)

$$SiCl_{4(v)} + 2H_{2(v)} \rightleftarrows Si_{(s)} + 4HCl_{(v)}$$

(3) (reduction)

$$GaCl_{(l)} + \tfrac{1}{2}H_{2(v)} + \tfrac{1}{4}As_{4(v)} \rightleftarrows GaAs_{(s)} + HCl_{(v)}$$

(4) (disproportionation)

$$2SiI_{2(v)} \rightleftarrows SiI_{4(v)} + Si_{(s)}.$$

With all of the systems the following criteria must be met. Reactive species must be transported at appropriate partial pressures to the substrate surface which must be heated to a sufficient temperature to initiate a heterogeneous reaction. One product of this reaction must be the material of the film semiconductor + dopant and any other products must be sufficiently volatile to be removed in the gas stream.

For pyrolysis and reduction reactions, a dynamic open-tube flow reactor operating at atmospheric pressure is normally used, and the basic elements of such a system are shown schematically in Fig. 8. Known partial pressures of reactant gases, including dopants, are introduced into a stream of hydrogen or inert carrier gas flowing at a pressure slightly greater than atmospheric through the reactor containing the heated substrates. In practice, multi-slice systems are frequently used, and for silicon technology in particular, up to one hundred 5 cm diameter substrates may be used in one run. They are heated indirectly by a silicon carbide coated graphite susceptor, which is itself heated by r.f. induction. In the case of GaAs growth, a furnace system is normally used for substrate heating.

One function of the carrier gas is to maintain a reasonably high linear gas velocity in the reactor, in order to obtain a uniform deposition rate of both semiconductor and dopant over the whole area of all of the substrates. The most frequently used carrier gas is highly purified hydrogen, and in reduction systems it therefore fulfils a dual purpose of transport and reaction, although the actual amount involved in the reaction is a very small fraction of the total flow.

The basis of the disproportionation system is the control of the equilibrium that exists between the higher and lower valency halides, mainly of silicon and germanium, illustrated by equation (4) above. The equilibrium is used to transport silicon from a high temperature zone to a lower temperature deposition zone, which contains the substrate, the whole process being carried out in a sealed tube at reduced pressure. This process is now only of historic interest, since although it was the method first used to prepare material for devices, the degree of control which can be exercised over layer thickness, perfection and impurity concentration is much less than in the flow systems, and it is no longer used technologically.

From these general principles it will be useful at this stage to consider more specifically two important processes used extensively in present day technology, one for silicon and one for GaAs. In this way the precise methods of growth and the characteristics of the grown material can be illustrated.

Dealing first with the preparation of silicon films by the hydrogen reduction of $SiCl_4$, a typical apparatus can be described by reference to Fig. 8. Between 20 and 100 substrates are mounted on the susceptor inside a rectangular sectioned silica reactor. The susceptor is heated by r.f. induction from a coil mounted outside the reactor. The first stage of the process involves *in situ* cleaning of the substrates by etching with HCl gas in hydrogen at a temperature of approximately 1200°C. This ensures that the grown layers will be of high crystallographic perfection, having few stacking faults and dislocations. $SiCl_4$ is then introduced into the highly purified hydrogen at a concentration of ~0·1 mole %. In a modern production process, the total gas flow rate would be > 100 litre min^{-1}, and the growth temperature 1100°–1200°C, producing a growth rate of ~1 µm min^{-1}. Layer thickness can readily be controlled to a reproducibility and uniformity of ±10%. Dopants are

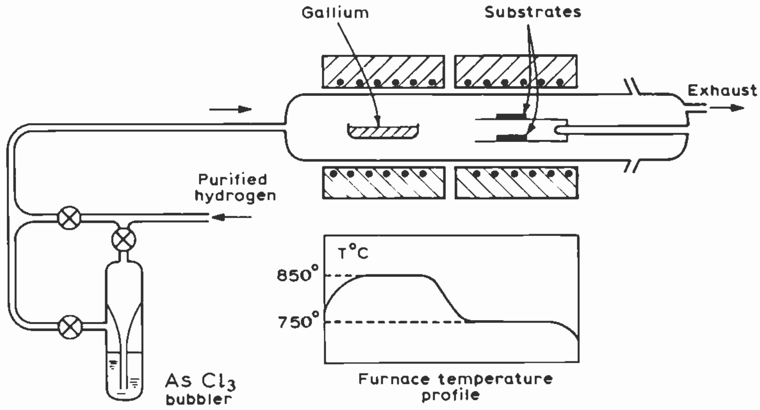


Fig. 9. Schematic diagram of $Ga/AsCl_3/H_2$ systems for the preparation of GaAs films.

introduced as the hydrides of Group III and V elements (B_2H_6, PH_3, AsH_3), and their concentration in the layers can be controlled over the range 10^{13}–10^{20} atoms cm^{-3} with a $\pm 10\%$ reproducibility except at the extremes of the range. In many production units this complete process is fully automated.

The preparation of autoepitaxial GaAs films is largely governed by the purity of starting materials, and the system most commonly used is based on gallium and arsenic trichloride (and hydrogen), since these are the purest sources of gallium and arsenic available. The apparatus is illustrated in Fig. 9. A two-zone furnace system is used to heat the quartz reactor, and again *in situ* gas phase etching is used to clean the substrates, five being the maximum number at one time in this system. Deposition is commenced by passing $AsCl_3$ in hydrogen over the gallium source at a temperature of about 850°C until it is saturated with arsenic. When saturation is complete, the substrate is inserted into the growth zone of the furnace, at a temperature of about 750°C. Continued passage of $AsCl_3$ over the arsenic saturated gallium produces a gas phase mixture of $AsCl_3$ and GaCl with hydrogen, and from this GaAs is deposited on the substrate. Gas flow rates are much lower than in the silicon system, being about 0·2 litre min^{-1}, with layer growth rates of about 0·3 µm min^{-1}. Doping is carried out with sulphur, tellurium, selenium or zinc, either in the form of the elemental vapour species or as hydrides: concentrations in the solid can be between 10^{13} and 10^{18} atoms cm^{-3}.

The deposition of GaAs films is much more difficult to control than is the case with silicon, and the reproducibility is not nearly so good. The principle reason for this is that many growth parameters are sensitive functions of the vapour phase ratio of gallium to arsenic species, which is not amenable to continuous control with existing technology.

3.4 *Autodoping*

One aspect of the chemical vapour deposition process which is common to all systems is the redistribution of impurities initially present in the substrate into the growing layer, a phenomenon normally referred to as autodoping. Redistribution can take place by two mechanisms (i) solid-state diffusion from substrate into the film during growth, which occurs whenever the substrate is more heavily doped than the film; (ii) out-diffusion of the dopant from all surfaces of the substrate into the gas phase, and subsequent reincorporation into the growing film.

Autodoping imposes a limitation on the control of the impurity concentration profile between layer and substrate, which always deviates from the ideal abrupt form. Although nothing can be done to reduce solid-state diffusion below the extent appropriate for a particular temperature and concentration gradient, out-diffusion can be restricted to the front surface of the substrate by masking all other surfaces with SiO_2 or high resistivity silicon.

3.5 *Growth from Elemental Sources*

Growth directly from the elements using vacuum evaporation or sublimation techniques has only been of academic interest, (see, for example, the review by Francombe and Johnson[27]) and has not featured at all in technological application. There are two main reasons for this: (i) it is much more difficult to control the dopant concentration than in chemical vapour deposition; (ii) to obtain films of high crystallographic perfection it is necessary to use ultra-high-vacuum techniques so that the substrate surface can be maintained free from contamination.

However, a modification of this technique, known as molecular beam epitaxy, is receiving considerable attention at the present time.[32] Molecular beams of Group III and Group V elements produced from Knudsen effusion cells impinge on a suitable substrate in a u.h.v. system. In this way precisely controlled amounts of the elements arrive at the substrate surface, affording accurate control of stoichiometry, doping and thickness, and since growth rates are low, ($\sim$ 0·03 µm min^{-1}), very thin films can be grown, e.g. for GaAs f.e.t.s. This method may well prove extremely valuable in the preparation of Group III–V compound films for special applications, in spite of the difficulties of implementation.

3.6 *Growth on Foreign Substrates*

Work on epitaxy has largely been concerned with silicon growth on insulating substrates (sapphire, i.e. single crystal α-alumina; and Mg-Al spinel, $MgAl_2O_4$) and has been reviewed by Filby and Nielsen.[33] By using insulating substrates it was hoped to obtain dielectric isolation and thus improve the high frequency performance of circuits by reducing the capacitance associated with p-n junction isolation. There are, however, two major unsolved problems in this technology. The first is the difficulty of obtaining films of adequate crystallographic perfection, and the second arises from the inclusion of aluminium from the substrates in the layers. This exhibits rather uncontrollable precipitation behaviour on any subsequent heat treatment for device processing, with consequent variations in carrier concentration. For these reasons, this technique has not fulfilled its early promise, and even if the problems were overcome, conventional technology has been so far improved in terms of material preparation for high frequency performance that the technique might find little use.

4 Conclusions

The knowledge of material preparation processes which has resulted from studies of semiconducting materials has been widely applied. It is, for example, doubtful whether crystals (for example, of ruby and yttrium aluminium garnet) for laser use could have been prepared without the background information that these studies provided. Similarly the thin layers used for the studies of magnetic bubbles are prepared by techniques which were initially developed for semiconductors.

In the last twenty-five years much has been achieved. In particular, the sizes, uniformity and degree of perfection of bulk crystals have all increased enormously; a whole new set of techniques for thin layer production have been evolved; and, possibly more importantly, the various techniques have been developed to the point at which they can reproducibly produce the required materials at a reasonable cost. This progress could not have occurred without considerable advances in the techniques of crystal assessment; for example, x-ray topography, infra-red microscopy and solid-source mass spectrometry.

It is difficult to predict trends over the next twenty-five years. However, it seems reasonable to expect that the costs of the various processes will decrease. In part the reduction of cost will come from automation. Other savings will be achieved by further integration of the materials preparation and device making processes. We can also expect some new techniques to emerge and some of the existing ones (e.g. molecular beam epitaxy) to be further developed. Changes in production technology (e.g. the use of ion implantation) may also require new material preparation techniques.

5 References

1. Laudise, R. A., 'The Growth of Single Crystals', (Prentice Hall, Englewood Cliffs, 1970).
2. Harper, C. A., 'Handbook of Materials and Processes for Electronics', (McGraw-Hill, New York, 1970).
3. Gilman, J. J., 'The Art and Science of Growing Crystals', (Wiley, New York, 1963).
4. Brice, J. C., 'The Growth of Crystals from Liquids', (North Holland, Amsterdam, 1973).
5. Brice, J. C., Roberts, J. A. and Smith, G. 'Mass spectrometric studies of impurities in gallium arsenide crystals', *J. Mater. Sci.*, **2**, p. 131, 1967.
6. Bridgman, P. W., 'Physical properties of single crystals of tungsten, antimony, bismuth, tellurium, cadmium, zinc and tin,' *Proc. Amer. Acad. Arts Sci.*, **60**, p. 305, 1925.
7. Burton, J. A., Prim, R. C. and Slichter, W. P., 'The distribution of solute in crystals grown from the melt', *J. Chem. Phys.*, **21**, p. 1987 and p. 1991, November 1953.
8. Hurle, D. T. J., 'Mechanisms of growth of metal single crystals from the melt', *Progr. Materials Sci.* **10**, No. 2, p. 79, 1962.
9. Brice, J. C. and Whiffin, P. A. C., 'The temperature distribution in pulled germanium crystals during growth', *Solid State Electronics (GB)*, **7**, p. 183, February 1964.
10. Dikoff, J. A. M., 'Cross-sectional resistivity variations in germanium single crystals', *Solid State Electronics*, **1**, p. 202, 1960.
11. Brice, J. C. 'Phase relationships in semiconductors', Chapter 3 in D. Shaw (Editor), 'Atomic Diffusion in Semiconductors', (London, Plenum, 1972).
12. Czochralski, J., 'Ein neues Verfahren zur Messung der Kristallizations-geschwindigkeit der Metalle', **92**, p. 219, 1918.
13. Gross, U. and Kersten, R., 'Automatic crystal pulling with optical diameter control using a laser beam', *J. Crystal Growth*, **15**, p. 85, 1972.
14. O'Kane, D. F., *et al.*, 'Infra-red tv system of computer controlled Czochralski crystal growth', *J. Crystal. Growth*, **13**, **14**, p. 624, 1972.
15. Gartner, K. J., *et al.*, 'An electronic device including a tv-system for, controlling the crystal diameter during Czochralski growth', *J. Crystal Growth*, **13**, **14**, p. 619, 1972.
16. Domey, K. E., 'Computer Controlled Growth of Single Crystal Ingots', IBM Technical Report No. TR ZZ.1190, 1971.
17. Patzner, E. J., *et al.*, 'Automatic diameter control of Czochralski crystals', *Semicond. Prod.*, p. 25, October 1967.
18. Brice, J. C., 'Analysis of the temperature distribution in pulled crystals', *J. Crystal Growth (Netherlands)*, **2**, p. 395, December 1968.
19. Brice, J. C., 'An analysis of the factors affecting dislocation distributions in pulled crystals of gallium arsenide', *J. Crystal Growth (Netherlands)*, **7**, p. 9, July 1970.
20. Theurer, H. C., US Patent No. 3,060,123, 1952.
21. Keck, P. H. and Golay, M. J. E., 'Crystallization of silicon from a floating liquid zone', *Phys. Rev.*, **89**, p. 1297, 15th March 1953.
22. Nelson, H., 'Epitaxial growth from the liquid state and its application to the fabrication of tunnel and laser diodes', *RCA Review*, **24**, p. 603, December 1963.
23. Pedersen, C. J. (ed.), 'Gallium Arsenide', Chapter 1, (Institute of Physics and Physical Society, London, 1968).
24. Small, M. B. and Barnes, J. F., 'The distribution of solvent in an unstirred melt under the conditions of crystal growth by liquid epitaxy and its effect on the rate of growth', *J. Crystal Growth (Netherlands)*, **5**, p. 9, Februaiy 1969.
25. Newman, R. C., 'A review of the growth and structure of thin films of germanium and silicon', *Microelectronics and Reliability*, **3**, p. 121, September 1964.
26. Joyce, B. A., 'Growth and perfection, of chemically deposited epitaxial layers of Si and GaAs', *J. Crystal Growth* (Netherlands), **3**, p. 43, 1968.
27. Francombe, M. H. and Johnson, J. E., 'The preparation and properties of semiconductor films', in 'Physics of Thin Films', Vol. 5 (G. Hass and E. R. Thun, eds), p. 143, (Academic Press, New York and London, 1969).
28. Schwartz, B., 'Epitaxial materials', in 'Vapour Deposition', (C. F. Powell, J. H. Oxley and J. M. Blocher, eds), p. 612, (Wiley, New York, 1966).
29. Feist, W. M., Steele, S. R. and Readey, D. W., 'The preparation of films by chemical vapour deposition', in 'Physics of Thin Films', Vol. 5, (G. Hass and E. R. Thun, eds), p. 237, (Academic Press, New York and London 1969).
30. Li, C. H., 'Epitaxial growth of silicon and germanium', *Phys. Status Solidi (Germany)*, **15**, No. 1, p. 3, and No. 2, p. 419, 1966.
31. Joyce, B. A., 'The chemical vapour deposition of single crystal films', in 'The Use of Thin Films in Physical Investigations', (J. C. Anderson, ed.), p. 87, (Academic Press, London 1966).
32. Cho, A. Y. and Hayashi, I., 'Epitaxy of silicon doped gallium arsenide by molecular beam method', *Met. Trans.*, **2**, p. 777, 1971.
33. Filby, J. D. and Nielsen, S. 'Single-crystal films of silicon on insulators', *Brit. J. Appl. Phys.*, **18**, p. 1357, October 1967.

Manuscript received by the Institution on 19th October 1972. (Paper No. 1492/CC 156.)

UDC 621.382.002

The technology of semiconductor manufacture

M. SMOLLETT,
Ph.D., B.Sc., A.R.C.S., D.I.C.*

SUMMARY

The basic technologies which have been developed to satisfy the production of semiconductor devices are outlined.

These technologies commenced with point-contact alloy and alloy-diffused techniques. Diffusion was then developed giving the means of making structures such as rectifiers, transistors and thyristors.

Silicon planar technology and the assembly methods used are then described as applied to discrete devices and integrated circuits (both bipolar and m.o.s.).

* *Semiconductor Division, Mullard Ltd., Southampton.*

1 Introduction

Early semiconductor technology was a primitive art using impure and polycrystalline materials with elementary pulsed point-contact methods for making diodes and the early transistors. Metallurgical and crystal growing knowledge was harnessed to making alloy junctions and then diffusion techniques coupled to photolithography gave more precision.

With the advent of 'planar' junctions the key to integrated circuits appeared, leading to the ability to manufacture sub-systems on chips of silicon.

This paper gives an idea of the main elements of semiconductor technology as they have progressed over the last twenty-five years.

2 Germanium Point Contact and Alloyed Technology

The point contact method of junction formation in germanium was the first practical technique used in the production of devices.

The junction is formed by the electrical pulsing of a whisker of metal such as tungsten or gold which is held in intimate contact with a germanium wafer. The majority of the stages of manufacture are on multi-position automatic machines from the envelope manufacture to the final sealing, measuring and finishing of the device. This old method of junction formation is still used today in this high-automated form to make millions of diodes. An example of the manufacturing equipment is shown in Fig. 1.

The alloy method of junction formation has been very widely used since the early days of germanium devices and indeed in the germanium power transistor field it is still the main production technique of today.

The basis of the technique is that if a pellet of indium is heated to about 500°C in contact with an n-type germanium wafer the pellet will spread and alloying into the germanium occurs. The depth of alloying is dependent on the temperature and the size of the indium pellet: if the germanium wafer has its faces parallel to the $\langle 111 \rangle$ plane, then a flat alloy front results. On subsequent cooling the germanium which is in solution in the molten indium recrystallizes on to the germanium wafer and since this layer is doped with indium a p^+-n junction is formed. By using two indium pellets, one on each side of the germanium wafer, p-n-p transistor structures are formed. The same basic techniques can be applied to form n-p-n structures except that p-type germanium and lead-antimony pellets are used.

In the high-frequency device field this simple alloy junction technique had many variants in the search for better control of the very small dimensions necessary for high-frequency performance. An illustration is the alloy-diffused technique where both pellets are alloyed from one side, one forming the emitter and the other the base, leaving the p-type germanium wafer as the collector. The principle of this technique is that if alloying of an n and p-doped pellet to germanium is carried out in the region of 800°C then in addition to the alloy junction formation there can also be a diffused junction formed

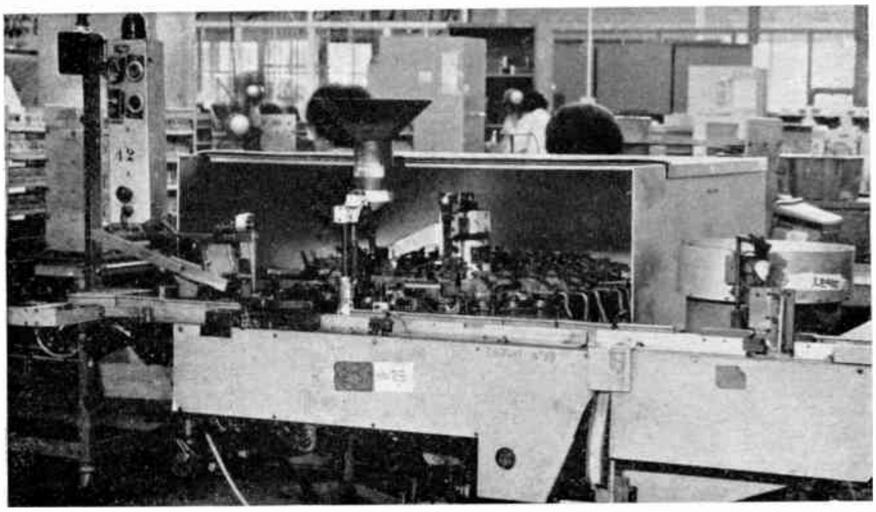


Fig. 1. An example of automatic diode making equipment.

by dopant diffusion from the alloy front. By choice of dopants and their levels in the pellets, transistors with very narrow base widths are attainable.

The usual method of assembling alloy devices is to alloy a metal tab to the germanium wafer, which together with contact wires are resistance welded to the header. The wires are then soft soldered to the alloy pellets. Before encapsulation the completed assembly is electrolytically etched to clean the exposed junction regions.

3 Silicon Alloying and Diffusion Technology

An alloy junction forming technique similar in principle to that used for germanium has been and still is used in the production of low power silicon diodes and transistors. The most successful alloying material to silicon is aluminium to form p^+-n diodes and p-n-p transistors. The process stages and the techniques themselves are very similar to those used in germanium: that is alloying in jigs, mounting by soft soldering and chemical etching of the final device.

Alloy junctions, in this case n^+-p junctions formed by alloying of gold-antimony to already diffused silicon wafers, were used for some time to form simultaneously the cathode junction and contact in thyristor manufacture.

However, the benefits in control, ease of manufacture, yield and the attainment of superior device characteristics of the diffusion technique for silicon, were very soon recognized and applied particularly to the mass production of rectifiers.

In the diffusion process the dopant is deposited on the silicon slice at a high temperature and is the source for diffusion into the silicon. The most generally used dopants are phosphorus and arsenic for n-type diffusions and boron and gallium for p-type diffusions.

There are two main methods of deposition on to the silicon slice: the sealed-capsule and the open-tube method. The sealed-capsule method relies on the transfer of elemental impurity from a source to the silicon slice in an evacuated and sealed tube. The open-tube method relies on the transfer of dopant in vapour from the furnace gas stream. Examples of this will be given later in the description of the planar process.

An important discovery in the early diffusion work was the realization that thermally-grown silicon oxide could form a barrier against diffusion. Junction areas could then be defined by removing oxide from the areas where diffusion was required.

Using the twin techniques of diffusion and oxide masking, device structures such as diodes, transistors, thyristors and triacs are manufactured. In the low power diode and transistor field the manufacturing techniques are now full planar, a process fully described later, but in the power device field the non-planar structures are still dominant. In general, as a result of fabrication techniques, high voltage junctions in rectifiers, thyristors and power transistors are left unprotected by oxide and therefore exposed. This arises, in the case of rectifiers, from the chemical etching necessary following the division of the diffused slice. In the case of some power transistors, the collector junction is defined by mesa etching using photoresist masking techniques.

Power devices normally have nickel contacts to which a 'soft' solder like lead-tin or a 'hard' solder such as gold-germanium are used to make connexion to headers and other contact tabs. Hard solders generally need the attachment of thermal expansion compensating plates made of molybdenum or tungsten to prevent bowing and cracking of the silicon wafer.

To ensure good surface stability, rubbers or glasses are applied to the junctions and high voltage devices often have bevelled edges to relieve the electric field at the surface. Power device envelopes are constructed so that one high current contact is attached to a stud or flange for mounting to a heat-sink. The remaining lead-outs pass through metal to glass or ceramic seals or are moulded in plastic. Examples are illustrated in Fig. 2.

4 Planar Technology

The advent of silicon planar technology in 1960 solved the problem of how to manufacture transistors and diodes in silicon with improved and reliable electrical charac-

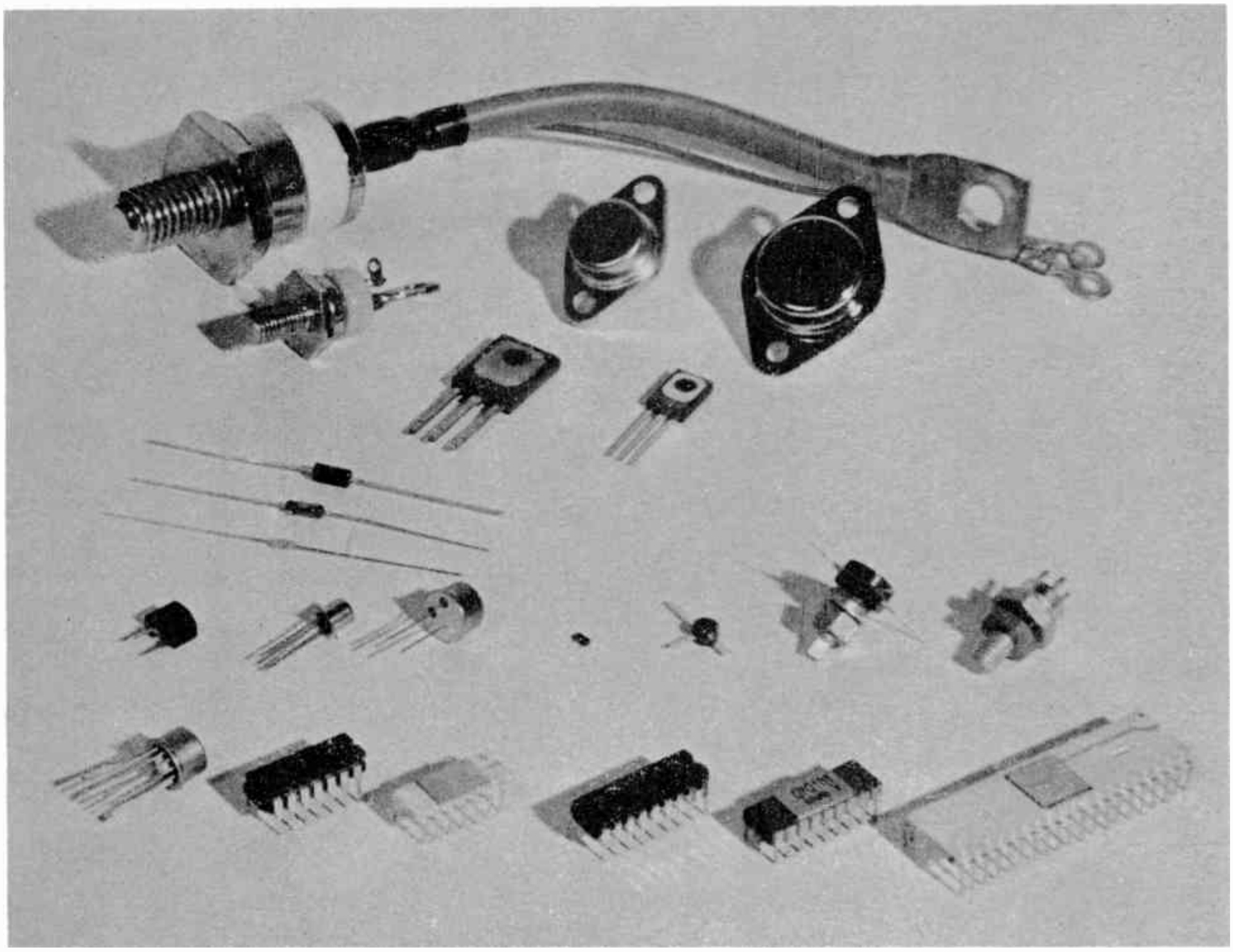


Fig. 2. Envelopes used for thyristors, power transistors, diodes, transistors, r.f. transistors and integrated circuits.

teristics at a cost in general less than the earlier germanium devices. As a result of this, the industry was set upon a path of massive re-equipment for this new technology. More refined transistor designs became possible with the precise planar techniques, solving the problems of how to reach high frequency, e.g. 1 GHz and also how to obtain power at frequency, e.g. 100 W at 500 MHz. A later consequence was the construction of the integrated circuit.

Planar technology resulted from the early research work on diffusion into silicon and the ability of the native silicon dioxide to mask against this diffusion. Silicon can easily be oxidized by high temperature furnace treatments in wet or dry oxygen. p-n junctions are formed by the diffusion of the appropriate impurity, usually boron or phosphorus, into the silicon and sources of both of these diffusants are readily available in tractable forms. Openings in the native oxide are achieved by photo-resist techniques, a technology which at the time was finding growing applications in such areas as printed circuit board manufacture and graticules. Planar processing technology is a continuous repetition of three basic sequences; oxidation to grow a masking oxide layer all over the slice: 'window' opening by photo-resist techniques; and the diffusion process to force impurities in to the exposed silicon in a controlled manner. This sequence can be followed by referring to Fig. 3 which shows the basic stages in the formation of a planar transistor of the early 'pear drop' design.

Four masks have to be designed and manufactured to complete the sequence outlined in Fig. 3. Mask making is one of the key technologies in planar manufacture, equally necessary with the simplest transistor as with the most complicated integrated circuit. The heart of the process revolves around the step-and-repeat camera, modern versions of which can cost in the region of £50,000. The shape of each stage, outlined in plan in Fig. 3, is cut in a peelable opaque film which is stripped in the required cut areas to reveal the mask detail as a transparent area. This transparent area will ultimately become the 'window' on the silicon slice after two stages of photographic reduction. The first stage, usually 20 : 1, reduces the detail onto a standard 50 mm × 50 mm (2 in × 2 in) photographic plate known then as a reticle. The step and repeat camera can possess up to six barrels, each barrel taking a stage reticle located accurately by means of pre-placed alignment marks. The step and repeat camera does two things, each barrel achieves a matched reduction of 10 : 1 and a stepping motion is imparted to one part of the camera, so that photoplates located under each barrel are exposed with a regular array of the original reticle design. Exposure

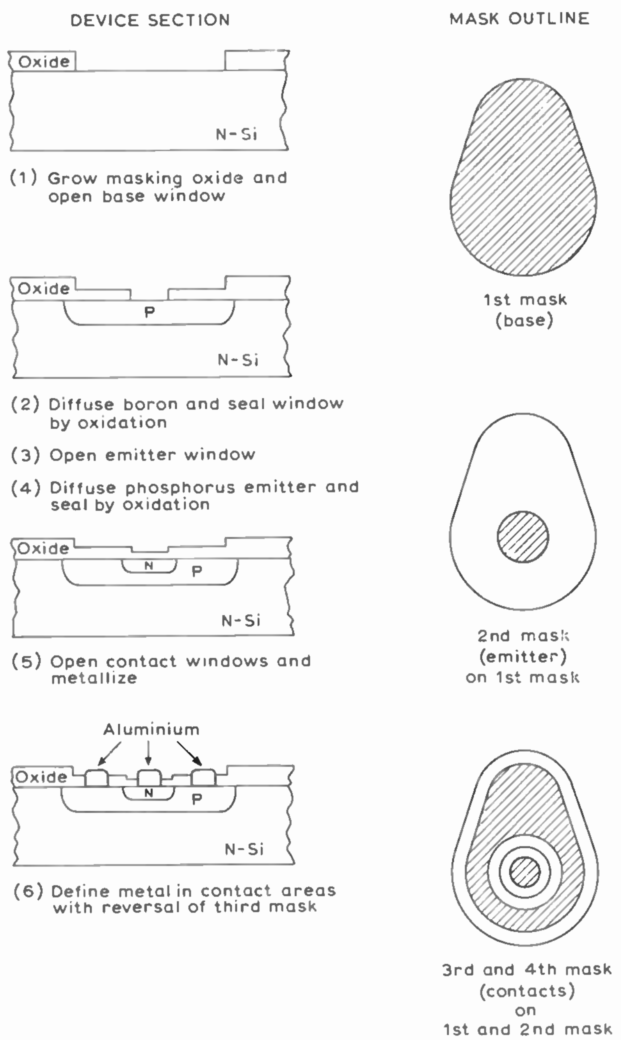


Fig. 3. Fabrication sequence for a 'pear-drop' planar transistor.

takes place by means of a high speed flashing light source. The accuracy of placement of each exposure is better than 1 μm.

The resulting plates from the camera are called masters and from these are produced a large collection of working copies as shown in Fig. 4. For complex integrated circuits, the cutting and peeling technique is extremely laborious and prone to human error, so here the design is coded in digital form onto a control tape which is then used to give instructions to a pattern generator. This equipment is similar to a single barrel step and repeat camera, except that its motion is not regular and the exposure is through an aperture which varies according to input instructions. The detail of the single stage reticle is built up in this equipment by exposing where required a series of overlapping blocks.

The quality of masks is paramount, in detail and dimensions, step and repeat pitch, and in the absence of defects. Manufacture takes place in a strictly controlled environment; temperature, humidity and dust level. Air conditioning plant and laminar flow conditions are required to achieve dust counts of the order of 35 000 0·5 μm particles/m^3 in rooms and below 3500 0·5 μm particles/m^3 in cabinets.

One of the important aspects of planar technology is the manufacture of many individual devices at one time. The process commences as the manufacture of numerous slices, each containing many hundreds of individual devices giving an enormous production capacity.

Historically the technique commenced using 2·5 cm (1 in) diameter slices, going later to 3·75 cm, then 5 cm, and the industry is now moving to 7·5 cm diameter slices. As the handling cost of a slice during manufacture is independent of size, the advantage of larger slices in terms of cheapness is obvious. The larger slices however make obsolete the equipment and tooling used for smaller slices right through from mask making to handling the finished slice. So during the life-time of the planar process, equipment has rapidly become obsolescent and each succeeding generation of equipment has become more powerful in terms of capacity and productivity, but in all cases considerably more costly so that high loading is needed to ensure profitability relative to depreciation.

The first stage in manufacture is oxidation. Here in modern equipment, in excess of 100 slices may be processed in wet oxygen at about 1100°C with the temperature controlled to better than ±1 deg C over a hot zone of 60 cm. A typical masking oxide grown is about 0·5 μm thick. Photo-resist technology follows. Individual slices are spun coated with a film of resist to a thickness of about 1 μm. This is dried by baking. The first mask is brought into contact with the slice and exposed. Development reveals the exact copy of the mask images all over the slice. Next, the layer remaining is hardened by further baking to resist the chemical etching used to dissolve away the unprotected oxide. After window etching, chemical removal of the resist prepares the slices for the following diffusion stage.

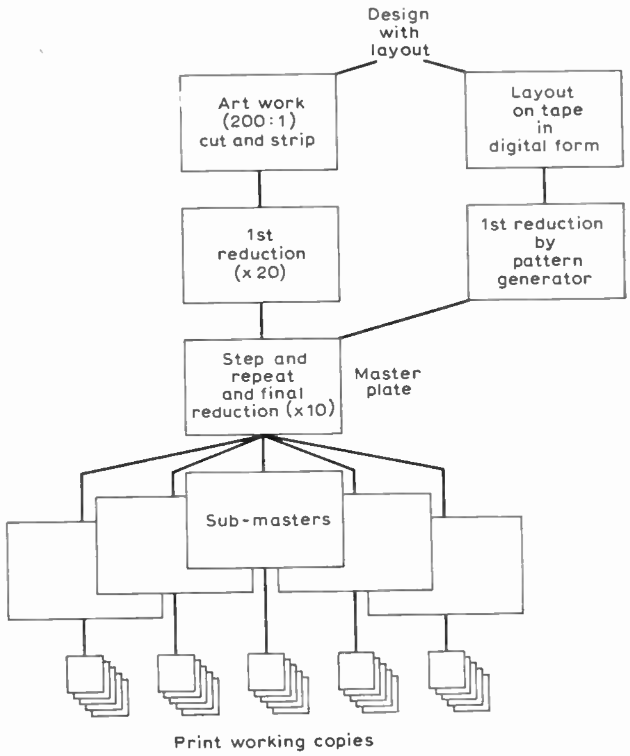


Fig. 4. Mask making sequence for production copies.

The diffusion processing unit contains furnaces set to each process condition, some for oxidation, some for impurity diffusion. Base diffusion requires two furnace-treatments to form the simple p-n junction diode. First there is a glass forming stage, where compounds of boron such as BBr_3 (a liquid) or B_2H_6 (a gas) are passed through the diffusion tube in the presence of oxygen. The boron compounds are decomposed and the slices become coated with a thin film of boron-rich glass. From this glass, boron diffuses into the silicon in the open windows. The glass is then removed chemically from the slice and submitted to further oxidation in the second furnace. This treatment grows the sealing oxide in the exposed windows and also the impurities are driven deeper into the silicon. Diffusion and oxidation are both very precise processes dependent on furnace gas conditions, temperature and process time. The schematic layout of a diffusion furnace is shown in Fig. 5.

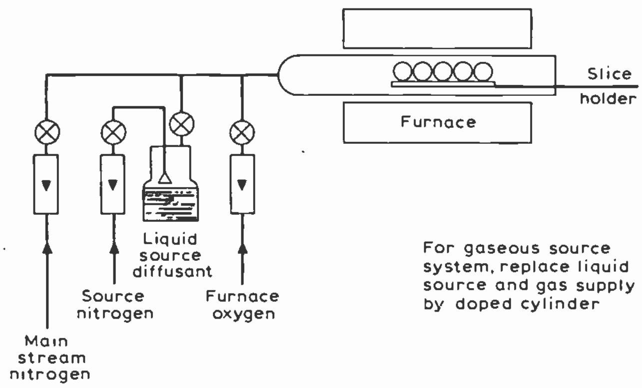


Fig. 5. Typical liquid source system for diffusion.

The same cycle is repeated with the emitter window using the second mask. Here it is important to realize that the emitter window is required to be opened in all patterns on the slice and correctly positioned with respect to the first mask. As described, the step-and-repeat camera accurately achieves this during mask manufacture; now the same accuracy is necessary in a superposition alignment. This is typically done manually, with the operator viewing not only both mask and slice together, but also, by means of a split-field microscope, two extremes of the slice. In this way, with the mask out of contact, the correct adjustments can be made in direction and rotation so that the superposition of all images occurs to the order of 1 μm. Exposure takes place with the mask and slice clamped in contact. Because the emulsion image on the mask is soft, it quickly becomes damaged during use and its life is limited to about 10 exposures. This explains the need for the manufacture of a large number of copies, all of which have to register with respect to each other.

In the very latest equipment alignment is completely automatic without the need for operator judgement and manipulation. Out-of-contact exposure is also possible. For example, one such technique is based on a projection lens system. Masks can also be generated by evaporation of a hard metal such as chromium with good wearing characteristics, or in a medium transparent to visible light but opaque to the ultra-violet used for exposure. The latter assists in difficult alignment situations and can be made from very thin layers of silicon or transparent metal oxides.

Returning to the planar sequence, the open emitter window is then subjected to a phosphorus diffusion, using either $POCl_3$, PBr_3 or PH_3 as a phosphorus source. Like boron, the deposition is a glass and the window is finally sealed by oxidation during the emitter drive-in process.

Because diffusion can be accurately controlled, the spacing between the boron base (p-type) junction and the phosphorus emitter (n-type) junction can be held to a variation of less than 0·1 μm. Between these junctions is the active base region of the transistor across which current transfer takes place.

The transistor has now been completely formed but is inaccessible because of the masking protective oxide. Here then is the great attribute of planar technology. The junctions are protected during manufacture and during subsequent electrical operation. The charge effects which have been shown to occur at the bare surface of semiconductors are minimized by the protective oxide giving more reproducible electrical characteristics and enhanced reliability.

Access to the transistor requires a third masking stage to open contact areas. The whole slice is then coated by evaporation with aluminium metal, about 1 μm thick, which is defined in the contact areas by the final mask and an appropriate etching step. The back surface of the slice will be used as the collector contact area. Great attention to process control is paid during manufacture, ensuring that oxide thicknesses and diffusion parameters are maintained. Photo-resist work is visually inspected for defects and dimensions. Cleanliness and correct handling are essential with environmental control through cleanrooms and frequently the more sensitive stages are equipped with individual laminar flow hoods for added protection against dust.

From now on, the final stages of manufacture are assembly operations; to place the individual device chip into a suitable package for protection against hostile environments and to allow electrical connexions to be easily made and power dissipated as heat to escape. These operations are expensive. To minimize costs, each individual transistor on the slice is contacted by probes and given a thorough electrical check before being passed to assembly. Probing stations are part of the test equipment and these can be set to index along rows of transistors on the slice, sense the edge of the slice and then return along the next row. Reject devices are marked during tests. Devices within specification and visually good are separated from the slice by scribing with a diamond stylus and cracking into individual chips. These chips are then mounted onto gold plated headers or frames by a soldering technique based on the eutectic reaction at about 400°C of gold in contact with silicon. This contact becomes the collector contact of the transistor. Because of the size of the contact areas on the chip, usually about 100 μm diameter, connexion is made with fine gold or aluminium wires of 25 μm diameter unless current carrying capacity requires something larger. In the peardrop transistor of Fig. 3 the contact areas are the emitter metallized centre and the enlarged offset base area on one side of the base contact ring.

Two techniques are available for joining the contact areas on the chip to the lead-out wires on the header. One uses gold wire fed through a capillary and is cut by a flame to form a ball. This ball is then positioned on the chip contact area and deformed under pressure applied through the capillary. During this operation the chip and header are held at about 300°C. The bond to the post is then made similarly using only the edge of the capillary as a wedge across the wire. The wire is then flamed off ready for the next bond leaving the tail which is subsequently removed. The alternative technique is to use aluminium wire and ultrasonic bonding. This is carried out cold. As before, the wire is positioned on the contact area and then scrubbed ultrasonically under pressure to effect the bond. In each case manipulation and dexterity is required, working with the aid of a microscope. The rates possible for this operation can typically be 30 bonds a minute. This type of equipment is shown in Fig. 6. In some cases this bonding process has been made automatic and free of manual skills.

Fig. 6. Device lead wire bonding in assembly.

As a final stage, the header is sealed with a cap to form an hermetic enclosure or, in the case of a grid-frame assembly, injection moulded into a suitable plastic. The package is tailored to the application of the device, special low inductance and capacitance headers being required for high-frequency operation and massive heat sinks for power operation (see Fig. 2).

The majority of transistors are formed on epitaxial substrates rather than homogeneous substrates. This is to reduce collector series resistance. A thin layer, 10–12 μm thick, of the required resistivity is grown onto the surface of a polished slice. This allows the polished substrate slice to be considerably lower in resistivity. Growth takes place in radio-frequency heated reactors involving the decomposition of silicon compounds such as silicon tetrachloride. During growth, impurities are added in a controlled manner to achieve the required resistivity.

The basic diffusion processes relying on boron and phosphorus enable n-p-n transistors to be more easily made than p-n-p types. Boron and phosphorus have similar diffusion constants but on oxidation behave differently. Phosphorus tends to remain in the slice or 'piles-up' whilst boron readily depletes into the growing oxide, a process called segregation. So p-n-p types require more specialized diffusion techniques. The same complication arises in the epitaxial substrate process. For n^+-n substrates besides phosphorus there are also arsenic and antimony, both of which are suitable doping impurities. The latter two are preferred because of their diffusion constants which are lower than phosphorus. With p^+-p substrates, gallium and aluminium diffuse much faster than boron and are unacceptable and even boron itself is not ideal. Diffusion constants lower than boron and phosphorus would be preferred to minimize out-diffusion from the high-low transition at the substrate-epitaxial boundary during device processing.

Mention must also be made of the various features built into manufacture to ensure reliability by designing out failure mechanisms. These range from the control of metallurgical situations which can occur during assembly due to the different metals being joined, to current carrying limitations caused by electromigration effects, which is metal transported by the high current densities possible in thin conducting films. Also silicon dioxide is not the perfect protective film and during device operation contaminant ions can migrate through the oxide under the presence of an electric field. Additional protection has to be provided by barrier layers or layers like phosphorus doped silicon oxide which gather and retain these mobile contaminants. These layers are necessary to ensure that the devices retain their required characteristics throughout their operating life.

5 Integrated Circuit Technology

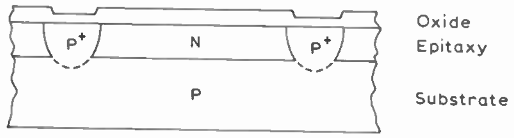
Integrated circuit technology is possible because in addition to transistors and diodes, resistors and capacitors may be made in silicon and then connected up with aluminium 'wiring'. There are two main classes of integrated circuit; bipolar and metal oxide semiconductor (m.o.s.).

In bipolar circuits resistors, diodes, diode-capacitors and bipolar transistors are used for making systems which are linear, digital or a combination of the two.

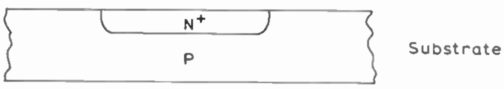
In m.o.s. circuits use is made of the fact that m.o.s. unipolar transistors are small, self-isolating and able also to act as loads. These circuits are almost entirely digital and form the first basis for large scale integration (l.s.i.).

5.1 *Bipolar Integrated Circuits*

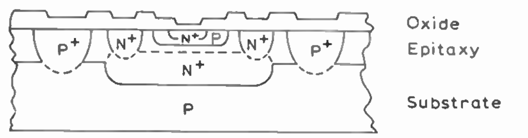
The basic planar processes already described for the oxidation of silicon, the opening of windows in the oxide using photomechanical techniques and the diffusion of the required n- or p-type impurity into the silicon surface



(a) Deep p^+-diffusion giving isolated n-type 'pockets'.



(b) n^+-buried layer diffusion in substrate slice.



(c) Collector wall contact to buried layer with an isolated n-type 'pocket'.

Fig. 7.

using the patterned oxide as a mask, can be used to make the circuit components like resistors, diodes and capacitors, as well as transistors.

Diodes require only one of the two diffusions necessary to form transistors.

Resistors are made by forming a structure that is really an elongated diode having a contact at each end. In operation the junction is reverse biased to minimize current flow across the p-n junction. The length l, between the two contacts divided by the breadth w, of the resistor multiplied by the sheet resistance ρ_s, of the diffusion determines the value of the resistance R, between the two contacts. Thus

$$R = \frac{\rho_s l}{w}$$

Since diodes, transistors, resistors and capacitors (the latter being provided by the capacitance of a reverse biased p^+-n junction) can all be made using the same planar processes, this leads to the manufacture of complete circuits, namely 'integrated circuits', in the silicon slice.

In order to be able to construct circuits with a flexibility approaching that afforded by wiring conventional discrete components, it is first necessary to be able to electrically-isolate individual devices from each other. This is achieved by surrounding the components with silicon of the opposite conductivity type and reverse biasing the p-n junction so formed. The only current flowing between adjacent components is then the negligible leakage current of the reverse biased junctions. The usual method used to form this structure is to start with a p-type substrate and grow an epitaxial layer of opposite conductivity type. Conventional planar techniques are then employed to carry out an isolating p^+ diffusion of sufficient depth to link up with the substrate material, forming a pattern of n-type 'pockets' in the original epitaxial layer into which the individual circuit components are subsequently made. This is illustrated in Fig. 7(a).

By using suitable mask designs and the normal planar diffusions for base and emitter regions used to manufacture discrete transistors, all the various components can be formed simultaneously. It then remains to interconnect these components into the required circuit configuration. This again can be accomplished using the same techniques as those employed to form electrical contacts to discrete devices. That is, a deposited and defined aluminium layer is used, but in this case it is employed not only to form electrical contacts but also the low-resistance interconnexions or wiring of the circuit on top of the oxide layer. As in the case of discrete devices many circuits are made on one slice of silicon; they are then separated into individual circuits

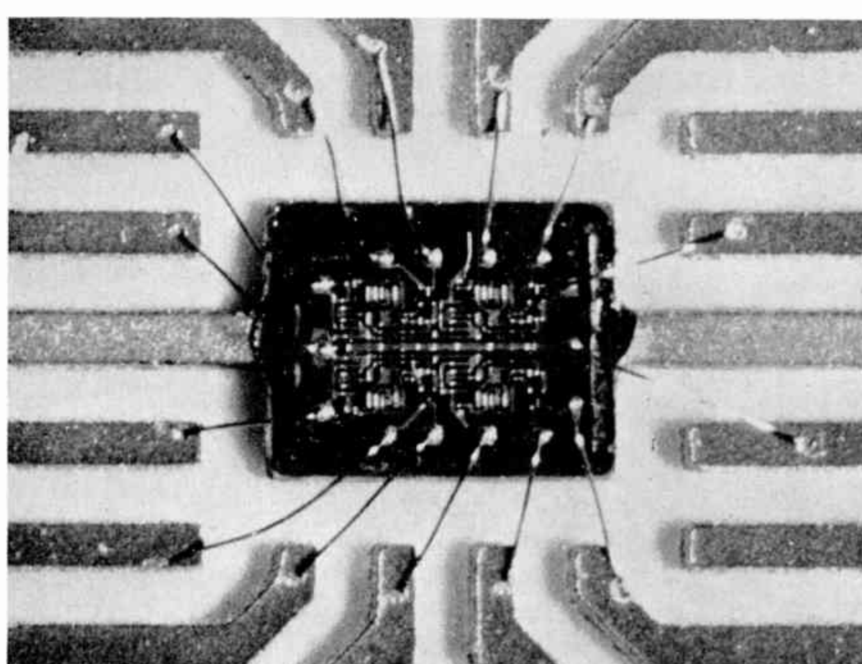


Fig. 8. View of an assembled integrated circuit on a grid-frame. (Chip size 1·5 × 1·1 mm).

by scribing with a diamond and cracking. A typical individual circuit already mounted and wired to an assembly grid-frame is shown in Fig. 8.

The main difference in the construction of an integrated transistor compared with its discrete counterpart is that the collector contact of the former must be brought out onto the top surface, as the electrically-isolated substrate is normally bonded directly to the header. In general, the topside collector contact results in collector series resistance, i.e. the resistance between the collector-base junction and the collector contact is considerably higher than that of the corresponding discrete device. This is clear from the difference in current paths through the high resistivity collector material in the two cases. To reduce this effect, an extra diffusion is usually included right at the beginning of processing to reduce the resistance under the collector region, as shown in Fig. 7(b). Before the epitaxial layer is grown the silicon is oxidized, windows are cut and an n^+ buried layer diffusion is carried out in the regions which lie directly underneath where the components are subsequently to be formed in the epitaxial layer. The oxide is then removed prior to epitaxy. Arsenic or antimony are normally used for this diffusion because the diffusion coefficients of these elements are an order of magnitude lower than that of phosphorus. Consequently there will be less out-diffusion into the epitaxial layer during subsequent high temperature processing, with a reduction in risk of detrimentally affecting the characteristics of the p-n junctions.

In some applications, where it is important to achieve as low a collector series resistance as possible, a 'collector wall' or 'collector sink' diffusion is also carried out to link up the collector contact with the buried layer region which itself diffuses some distance up into the epitaxial layer during this process. This diffusion is similar in type to that of the n-type emitter but is carried out earlier in the process so that a deeper diffusion can be achieved before the shallow diffused transistor structures are formed. This contruction is shown in Fig. 7(c).

To reduce carrier lifetime and achieve higher switching speeds gold is employed to act as a lifetime killer. The gold is evaporated on the back of the slice and diffused in during the phosphorus emitter drive-in stage. Its high diffusion coefficient allows it to diffuse through the slice whilst the shallow emitter is being formed. Schottky clamping diodes across the collector to base junctions of the relevant transistors are an alternative method to achieve fast switching speeds. The diodes prevent the transistors going into heavy saturation. Aluminium is commonly used to form these diodes at the same time as the aluminium interconnect layer is made.

To further reduce the area occupied by the circuit, techniques are in evolution which

(i) reduce the width of the aluminium interconnect;

(ii) produce a two-level interconnect where one aluminium layer can cross another separated by a silicon oxide insulator;

(iii) reduce the area required for isolation, by for instance using oxide isolation instead of diffusion.

(This oxide isolation method is known as LOCOS and uses patterned silicon nitride to achieve the local oxidation of silicon.[1])

5.2 *M.O.S. Integrated Circuits*

Metal oxide semiconductor transistors (m.o.s.t.) differ from bipolar transistors in that they are majority carrier devices. The device action comes from the establishment of a conducting path between two separate diffused regions, source and drain, by the influence of a controlling gate electrode spanning both diffusions. It is therefore a field effect device and in this situation the electric field across the gate dielectric approaches 10^5–10^6 volts/cm. The gate dielectric is thin and usually of silicon dioxide or silicon nitride. The most common device structure is the p-m.o.s.t. formed with boron diffusion and operating with a negative voltage on the gate to induce positive charge in the underlying silicon thereby completing the conduction path between diffusions.

Although individual m.o.s. transistors are made as single gate triodes, or as double gate tetrodes, the major impact of this technology is in the digital field where the smallness of size and the use of m.o.s. loads permits large scale integration to be easily achieved. Here many thousands of transistors can be combined on a chip to form a switching system.

The critical manufacturing steps in m.o.s. technology relate to the control of oxide charge and oxide quality. This requires some explanation. When silicon is oxidized the transition from pure oxide to pure silicon is not abrupt but gradual and with this region is associated the positive charge which by experiment is found to exist in the oxide near to the silicon surface. The magnitude of this charge depends critically on the oxide growth conditions and the incorporation of annealing steps in the processing conditions. Other problems can exist due to the presence of mobile charges in the gate oxide or to polarization effects due to the presence of phosphorus gettering glasses. In double dielectric structures minute differences in conduction can build up charged interfaces. All of these features have to be controlled in manufacture. This control stems originally from the basic understanding derived from initial research, with confirmation obtained by repeated processing in production, combined with rigid quality control procedures. In m.o.s. these control techniques include the use of capacitance-voltage plots on capacitors formed from oxides used in device fabrication, especially after treatments at high temperature (300°C) during which different polarity bias voltages are applied to the capacitors.

In common with other device technologies, there are numerous m.o.s. manufacturing processes of which only two will be outlined. These are the two processes that are perhaps nearest to being accepted as standards: aluminium gate for high threshold ($V_T = 3{\cdot}0$ V) and silicon gate for low threshold ($V_T = 1{\cdot}5$ V). Each process can be manufactured with single polarity transistors either as p-channel equivalent to p-n-p bipolar transistors or n-channel equivalent to n-p-n transistors.

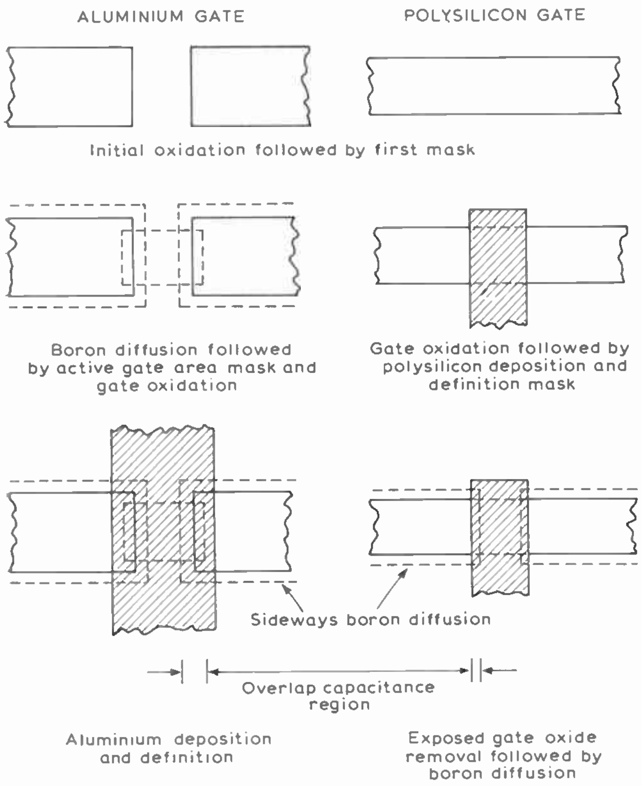


Fig. 9. Outline plan of aluminium gate and silicon gate m.o.s.t.s.

Alternatively both polarity types can be combined on a single chip to form complementary m.o.s. This is achieved by first preparing the n-type substrate with a p-diffused 'well' or 'pocket'. It is in this p-diffusion that the n-channel transistors are formed, with the p-channel devices formed in the undisturbed substrate.

As in bipolar technology, one transistor type was at the outset easier to make, this is the p-channel device using an n-type substrate, simply because the built-in oxide charge is positive. The simplest p-channel process is the aluminium gate process and requires four masking stages. First, as in all planar techniques, an initial masking oxide is grown in which for p-channel is opened the regions for boron diffusion. After boron deposition a thick oxide is produced approximately 1·5 µm thick, and the boron diffused further into the slice. This oxide is the isolating field oxide and an aluminium conductor crossing two parallel unrelated boron conductors must not form a parasitic m.o.s. structure. The active m.o.s.t.s are formed by a further photomechanical stage which opens windows in the field oxide to become either gates or contact areas. Into these regions the gate oxide is grown, typically only 1200 Å thick. The third mask repeats the window opening in the contact areas only, so that when the aluminium is deposited and defined by the fourth mask, conductor runs are formed terminating in both active gates and contacts to boron. The gate oxide and the aluminium gate metal must be 'clean'. Sometimes phosphorus gettering layers or silicon nitride barrier layers are incorporated in the gate dielectric under the metal. In bipolar technology filament evaporation of aluminium is normal, in m.o.s. to achieve very pure metal, electron beam evaporation techniques are preferred.

In the aluminium gate process therefore, the circuit can be traced at two levels; in the substrate as diffused boron which is self-isolating and in the aluminium as a conventional conductor. The transition from field oxide (1·5 µm) to active gate (0·12 µm) is a deep step requiring special understanding of the techniques to achieve aluminium conductor continuity over these steps. The plan of the aluminium gate m.o.s. is shown in Fig. 9.

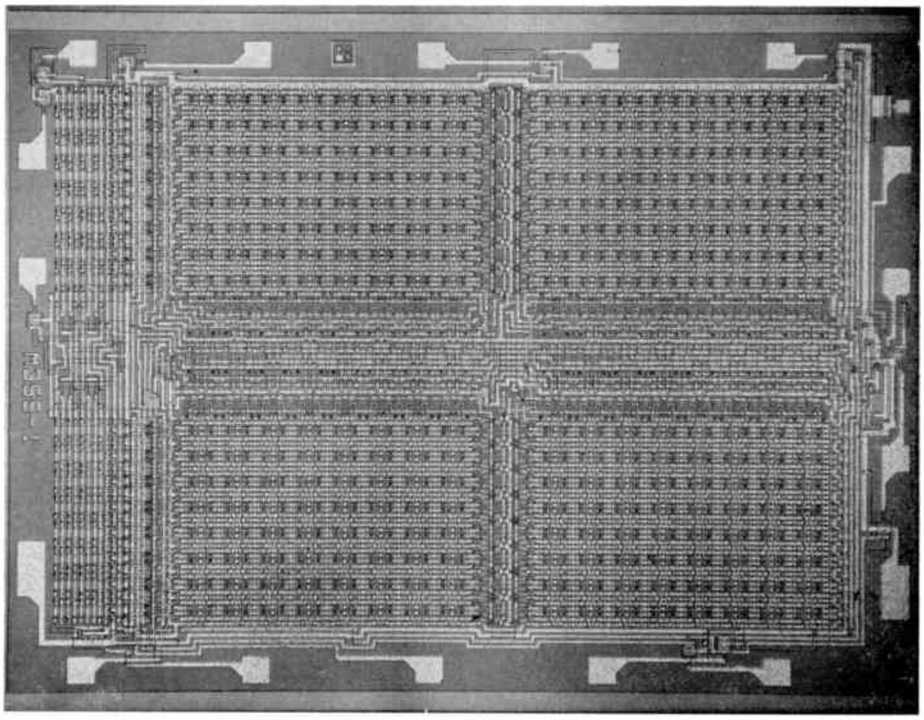


Fig. 10. A 1024-bit random access memory fabricated in p-channel gate m.o.s. (Chip size 4 mm × 3 mm)

As the process description indicates, two mask stages are required to superimpose the active gate area over the diffusions and the gate metal over the active gate area. These require alignment tolerances to be included in design to achieve the required overlaps which in themselves lead to unwanted feedthrough capacitances as well as consuming space.

The silicon gate p-channel process, apart from being a low threshold process, reduces the overlap problem because it is a self-aligned process. As its name implies, this process uses polycrystalline silicon to form the gate regions. The first stage is the growth of a thermal oxide approximately 1 μm thick. Into this oxide is cut, using the first mask, a window to contain boron diffusion areas plus the active gate areas. A gate oxide is then grown in this window, about 1200 Å thick, followed by polysilicon deposition all over the slice. The gates and conductors are then defined in this polysilicon layer by the second mask. By etching away the exposed gate oxide in the wells formed by the first mask the slices are prepared for boron diffusion with the gates already in position. Diffusion takes place into the substrate and polysilicon and sideways under the polysilicon at the source and drain edges. In this way the gates are automatically registered as shown in Fig. 9. The fabrication is then completed by a deposited dielectric layer over which the aluminium conductors run. Silicon gate can also be employed in n-channel technology, but there the diffusant is phosphorus into a p-type substrate. In n-channel technology the oxide charge has to be maintained at its minimum level and ⟨100⟩ orientation material is employed in place of the normally used ⟨111⟩.

A circuit typical of current technology in silicon gate (p-channel) as shown in Fig. 10. This shows a 1024-bit dynamic random access memory circuit containing over 4000 transistors on a chip of 12 mm^2.

5.3 *Encapsulation and Testing*

Integrated circuits (which for l.s.i. are between 10 and 36 mm^2 in size) are pretested on the slice then cut up, bonded and packaged in a similar way to planar transistors. Either ceramic or plastic dual-in-line packs are used (see Fig. 2) with usually 14, 16, 24 or 40 leads.

Testing of integrated circuits which may contain more than 10,000 transistor gates is very complex. Modern multi-probe testers costing over £100 000 are required to test what is a complete system. Several million tests must be made in a few seconds.

6 Future Technology

Although enormous progress has been made in developing semiconductor technology in the past 25 years, we may expect that the future will hold further refinements and changes. Ion implantation of the dopant will be used for greater control and registration.[2] Electron beam techniques could be the answer for producing finer and more dense geometries on the slice. Isolation techniques will be improved and completely insulated substrates introduced. Large scale integrated circuits will require larger computers for their design and test. Future semiconductor technology will allow us to design more and more electronics on smaller pieces of silicon; in particular multi-level techniques will become commonplace. Conventional device assembly as practised today will largely be replaced by complete sub-system modules, in effect miniaturizing the printed circuit board.

7 References and Bibliography

1. Kooi, E. *et al.*, 'LOCOS devices', *Philips Research Reports*, **26**, 166–180, June 1971.
2. Stephen, J., 'Ion implantation in semiconductor technology', *The Radio and Electronic Engineer*, **42**, pp. 265–83, June 1972.
3. Hunter, L. P. (editor), 'Handbook of Semiconductor Electronics' (McGraw-Hill, New York, 1956).
4. Genty, F. E. *et al.*, 'Semiconductor Controlled Rectifiers, Principles and Applications of p-n-p-n Devices' (Prentice-Hall, Englewood Cliffs, New Jersey, 1964).
5. Warner, R. M. (editor), 'Integrated Circuits, Design Principles and Fabrication' (Motorola Series in Solid State Electronics), (Mc-Graw-Hill, New York, 1965).
6. Grove, A. S., 'Physics and Technology of Semiconductor Devices' (J. Wiley and Sons, New York, 1967).
7. Burger, R. M. and Donovan, R. P., 'Fundamentals of Silicon Integrated Device Technology' (Prentice-Hall, Englewood Cliffs, New Jersey, 1967).
8. Beale, J. R. A., Emms, E. T. and Hilbourne, R. A., 'Microelectronics' (Taylor & Francis, London, 1971).

Manuscript received by the Institution on 27th October 1972. (Paper No. 1493/CC157.)

UDC 621.382.04977

Integrated circuit development

J. S. BROTHERS,
C.Eng., M.I.E.R.E.*

SUMMARY

The paper looks at the development of integrated circuits from its conception in 1952, through to the present multi-million dollar business. Major developments in both circuitry and process technology are highlighted and some indication is given of the direction in which the industry may move in the future.

* *Allen Clark Research Centre, The Plessey Company Ltd., Caswell, Towcester, Northamptonshire.*

1 Introduction

It is now some twenty years since the idea of an integrated circuit was first publicly announced. In those years, a considerable investment has been needed to solve the continuum of problems that emerged in taking the technologies through research and development, and finally into production.

There were periods, as in the early 1960s, when many considered that, because of yield difficulties, integrated circuits as we know them now would never be a commercial practicality. Even as late as 1964 protagonists of the hybrid circuit contended that their solution would be the most cost-effective. It is a testament to the success of the industry that what was earlier known as the monolithic integrated circuit to distinguish it from the hybrid variety, is now known simply as the integrated circuit, or i.c.

The variety of circuits now available exceeds even the optimistic forecasts of ten years ago. A major manufacturer can now have an inventory of over 2000 standard i.c.s, together with hundreds of 'specials'. This does not mean, however, that the user can yet find an integrated circuit for every application. The spectrum of uses covered by i.c.s is by no means complete but it is widening continuously. Uses of integrated circuits now range from spacecraft to wristwatches and computers to washing machines.

Early integrated circuits were a direct copy of their discrete component counterparts. It was only when a new generation of designers, who understood both circuits and the limitations of the technology, came to the fore that integrated circuits moved ahead. This occurred in about 1964. Such designers exploited the virtues of i.c. technology and overcame the restrictions it imposed by clever circuit design.

The complexity of integrated circuits has grown from three transistors and one resistor in 1960, through some 40 odd components in 1964, to some 12 000 components in 1972. It is this dramatic increase in complexity that has enabled the i.c. manufacturer consistently to meet his price predictions, but this move towards increased complexity has caused its share of problems. The rapid pace that the industry has maintained caused some users to delay placing production orders in the hope that new developments would reduce their costs. This attitude made it difficult for the i.c. manufacturers to recoup their investment at each stage of development, and it is a sad fact that few have produced a satisfactory return on the capital invested. The main beneficiaries of the boom in integrated circuits have undoubtedly been the system manufacturer, and the consumer.

In the space provided by a single paper it is impossible to do justice to the 100 or more significant contributions that have been made to integrated circuits. The author therefore has concentrated on the contributions he considers to be of greatest import, and apologizes for any significant omissions.

2 The Early Years

It is doubtful if anyone can determine a single point of origin for the field now called integrated circuits. Its evolution naturally progressed out of urgent need from many directions. However, in a paper presented in Washington in May 1952, G. W. A. Dummer, then of RRE, predicted: 'With the advent of the transistor and the work in semiconductors generally, it now seems possible to envisage electronic equipment in a solid block with no connecting wires. The block may consist of insulating, conducting, rectifying and amplifying materials, the electrical functions being connected directly by cutting out areas of the various layers.' Under contract to RRE the Plessey Research Laboratory at Caswell began work in 1957 to develop 'solid circuits' as they were then known. Whilst the original circuits bear little relationship to present day integrated devices, reports dating from 1959 onwards show considerable foresight in manufacturing techniques.

During this early period, considerable controversy raged as to whether thin-film or monolithic integrated circuits would prove the better commercially. Protagonists of the thin-film approach insisted that it would be impossible to obtain satisfactory yields on circuits where it was impossible to change faulty transistors.

In the USA the success of integrated circuits was ensured by the enormous sums of money made available to the manufacturers via the space and defence programmes. This money was available not only for research (as was the case in the UK) but also for volume production. This policy was reflected in the fact that the five major producers had a production rate of 5 million circuits a year by the end of 1963.

2.1 *Technologies*

Early integrated circuits, up to late 1961, were based on the 'mesa' transistor technology. A circuit would typically consist of three or four transistors with isolation between components being effected by a high resistivity silicon substrate. The transistors used, because of the limitations imposed by the technology, were inferior to their contemporary discrete equivalents.

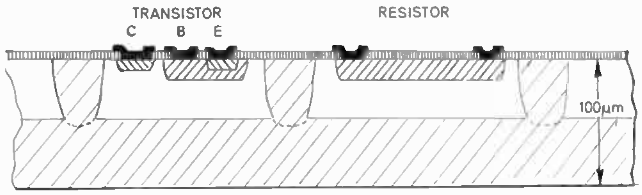
With the invention of 'planar' technology by J. A. Hoerni, of Fairchild, in March 1962, the development of integrated circuits was speeded considerably. The planar technology was developed rapidly to give p-n junction isolated transistors within the same semiconductor substrate. Concurrently with this, improvements in the photolithographic techniques used reduced the sizes of transistors considerably and enabled multiple circuits to be formed on one substrate by a process known as 'step and repeat'. Cross-sections of two of these first generation integrated circuit processes are shown in Fig. 1.

The industry was troubled considerably by yield in these early years, and this, together with the threat from thin-film circuits, caused some manufacturers to pursue circuits with built-in redundancy; the components required for the circuit were then selected and an interconnexion pattern applied. This approach enabled design cycles to be reduced at the cost of expensive silicon surface area. Such a technique as this which received some commercial support was the Texas Instruments 'Master Slice'.

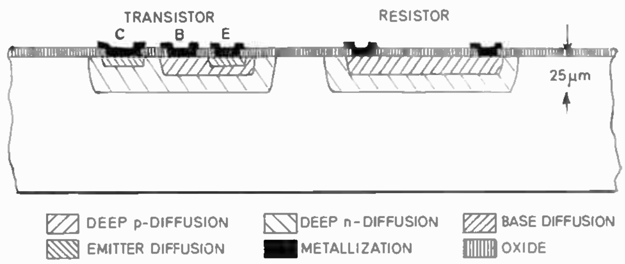
By adding more diffusions some manufacturers produced circuits with both n-p-n and p-n-p, but their commercial impact was small by comparison with triple-diffusion processes shown in Fig. 1. Some of these processes did contain, however, a technology improvement called epitaxy which was to cause the next major process leap.

2.2 *Circuits*

It is often wrongly assumed that all early integrated circuits were digital. Considerable interest was shown in analogue circuits, but this was not built upon because of the inferior characteristics of the transistors used. An amplifier with a gain-bandwidth product of 300 MHz was offered for sale.[1]



(a) The deep diffused process. The deep diffused integrated circuit begins with an n-type silicon substrate. A p-diffusion from both sides of the wafer is then used to create n-type islands. Transistor base and resistor regions are then p-diffused from the top and an n-type emitter and collector diffusion completes the transistor. Regrown oxide is etched away at required contact points and aluminium connexions are made.



(b) The triple diffused process begins with a p-type silicon substrate. A deep n-type diffusion next forms the isolated regions. Transistors and resistors are then constructed as in Fig. 1(a).

Fig. 1.

The industry at this time had not learnt to live with the restrictions imposed by its technology. Considerable weight was attached to the disadvantages in design and little to the advantages.[2] As is usual in the evaluation of any new technology, 'different' had become a synonym of 'worse'.

Commercial availability of circuits prior to the invention of 'planar' was poor. However, by early 1963 some 20 manufacturers were offering digital circuits for sale and eight were offering analogue circuits.

Digital circuits offered included,
- resistor-transistor logic (r.t.l.)
- diode-transistor logic (d.t.l.)
- resistor-capacitor-coupled logic (r.c.t.l.)
- transistor-transistor logic (t.t.l.)
- direct-coupled-transistor logic (d.c.t.l.)
- and
- emitter-coupled logic (e.c.l.).

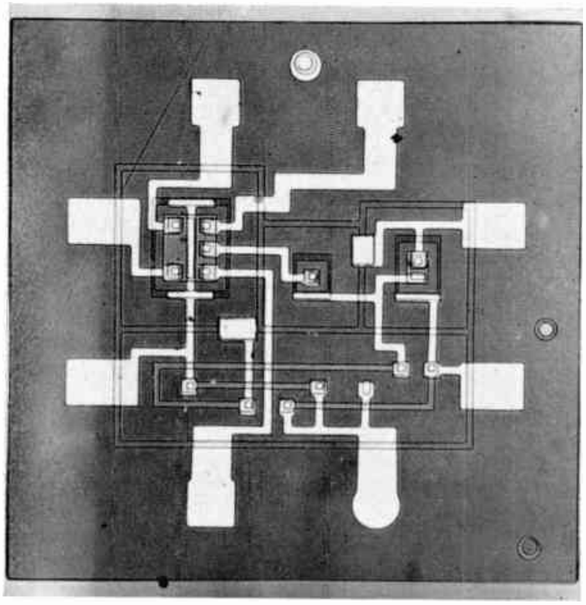


Fig. 2. A d.t.l. circuit (1963). An early 5-input d.t.l. NOR gate produced on a deep diffused process. Note the large separation between isolation diffusion and active devices (1963). (1·87 × 1·87 mm).

The t.t.l. available at this time was inferior to that used currently in that it had poor noise immunity and fan-out capability.[3] These circuits, as do present t.t.l. circuits, rely on the invention of the multiple emitter transistor by P. M. Thompson in 1961.[4]

It is of interest to note that although both circuits and processes were not fully developed digital switching speeds were high. The Motorola e.c.l. circuits were achieving 6 ns propagation delays.

Figure 2 shows a photomicrograph of an early d.t.l. circuit.

3 Integrated Circuits 'Come of Age'

During the early years, a considerable amount had been heard about the miracle of microelectronics, but until 1964 little had been shown. Aggressive marketing became a dominant trend and this showed in the pricing policies of the major manufacturers. One such manufacturer reduced his price to $12 a package in 1964, whilst at the same time stating that the I.C. Department did not expect to show a profit until 1965. Any would-be user of integrated devices was showered with leaflets and data sheets free of charge.

Integrated circuits had become a business. What was once an art was rapidly becoming a science; the green-fingered do-all was replaced by the specialist. Every aspect of the production process was thoroughly documented and where possible human operations were removed.

It was no longer sufficient just to be able to produce integrated circuits to sell them. The product had to be cost-competitive and reliable.

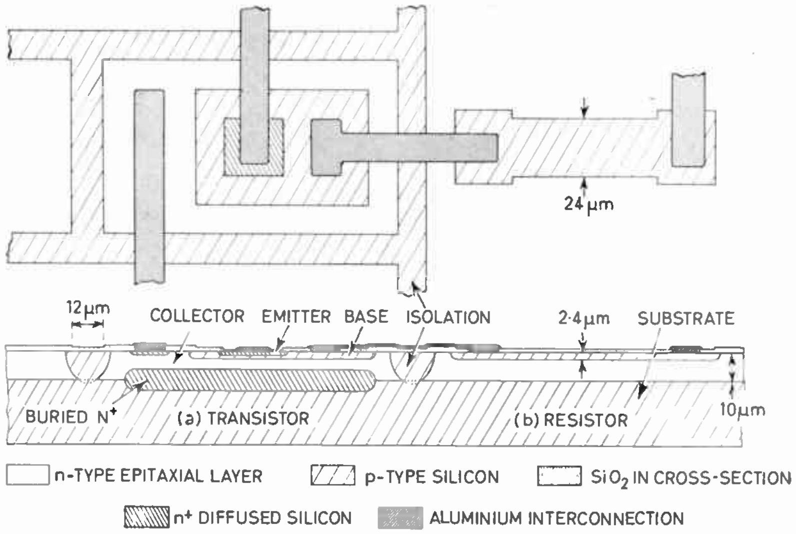
3.1 *Technologies*

The epitaxial technique referred to above in combination with a new development called a buried layer diffusion, depicted in Fig. 3, formed the basis for the new technologies. These second generation processes, with a few refinements, form the basis of most of today's production processes.

A period of consolidation then occurred where attention was turned to the various techniques employed in the fabrication of an integrated circuit. All the facets of the processing were analysed and improvements sought. The photolithographic process was one such process to undergo substantial change, the materials used were greatly refined and considerable benefits in yield were obtained. Such were the benefits of this approach that costs reduced by a factor of ten within about three years. The catch phrase of the time was 'Yields are so good now that we are expanding into a smaller factory.' This yield improvement was due not only to 'defect-free' processing but also to considerable reductions in the size of components.

A new radical development occurred in 1964, the introduction of the insulated gate field effect transistor in an integrated circuit process. The process, since it used silicon dioxide as the insulating medium, was called the metal-oxide-silicon (m.o.s.) process. As the techniques

Fig. 3. The buried layer process. The process commenced with a p-type substrate into which an n-type layer was diffused where transistors were to be firmed. A thin epitaxial layer was then grown (~10 μm). A p-type diffusion from the surface next formed the isolated regions. Transistor diffusions were then carried out as before but with reduced diffusion times to produce shallower junctions and transistors with higher operating frequencies.



for manufacture were similar to those for bipolar integrated circuits, devices acquired production status rapidly. The major advantage of m.o.s. was its simplicity of process and the compactness of its transistors. But announcement of m.o.s. was premature, with initial devices having reliability problems, and the commercial introduction of m.o.s. was delayed some two years.

Amongst other processes being developed at this time were oxide isolation, complementary m.o.s. and silicon-on-sapphire.

3.2 *Circuits*

With the introduction of buried layer processing, the development of analogue circuits took on a new life. Transistor characteristics were now comparable to those used in discrete circuits. The first truly commercial circuit was the operational amplifier and the first of these to achieve international acclaim was the μA702 designed by Widlar,[5] then at Fairchild. A circuit diagram of the amplifier is shown in Fig. 4.

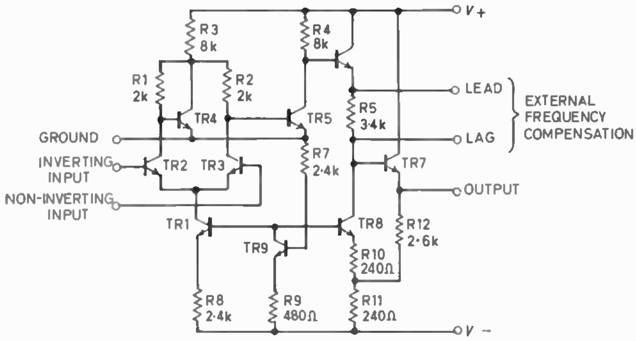


Fig. 4. An early operational amplifier (1964).

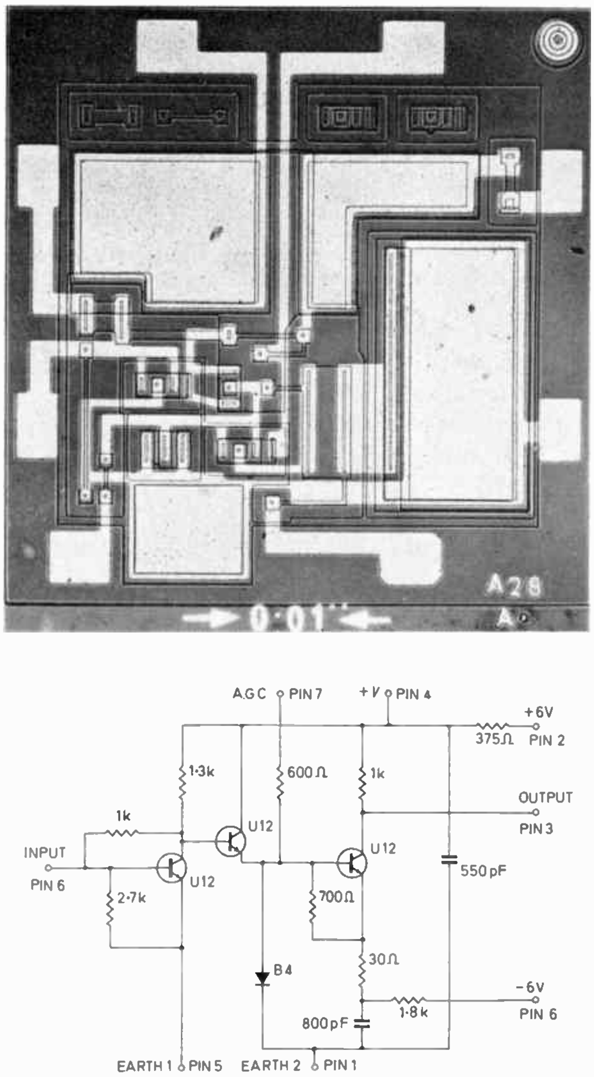


Fig. 5. An early wideband monolithic i.c. amplifier (1964). Chip photograph and circuit diagram of possibly the first wideband monolithic amplifier. Produced on an early buried layer process it had a bandwidth in excess of 140 MHz and a gain of 26 dB (1964). (Chip size 1·5 mm × 1·6 mm).

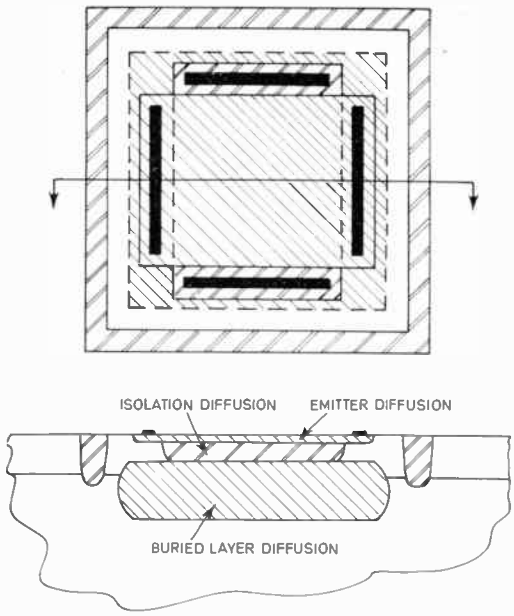


Fig. 6. Capacitor. Cross-section and plan view of a monolithic integrated circuit capacitor made by using a sandwich technique. Capacitance values of 3·5 pF/sq. thou. were obtained by connecting two junctions in parallel.

In Britain, in 1965, emphasis was being turned to exploiting what had previously been thought of as the deficiences of integrated circuit parasitics. It was realized by some designers that although these parasitic capacitances were possibly larger than those found in discrete circuits, their value was much more predictable, and, by using thorough design, these effects could be nullified. It was this knowledge that permitted M. J. Gay[6] to design what was to be the world's first i.c. wideband amplifier. The circuit diagram and chip photograph are shown in Fig. 5. Although this circuit had been fabricated and some samples sold, a paper on this subject written for a leading US Circuits Conference was refused on the grounds that the circuit was beyond the present 'state of the art'. This circuit owed much to a novel integrated circuit capacitor technique[7] developed at this time and is shown in Fig. 6.

The battle for supremacy in digital circuits was contested with considerable vigour, r.t.l., d.t.l. and t.t.l. all seeming likely winners at some stage. The contest was finally resolved in the favour of t.t.l. by virtue of strong marketing, lower costs and superior 'specsmanship'. Figure 7 shows a typical four-input NAND gate of this period.

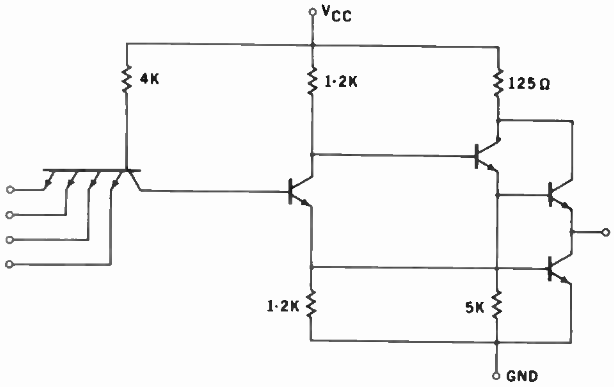


Fig. 7. An early production t.t.l. circuit (1964).

A major circuit development of this time was the provision of logic families for users. For the first time fully-compatible logic families of up to some twenty different functions became available. The families contained basic gates, various flip-flops, and more complex functions such as adder circuits. Complexity of these circuits ranged between 10 and 40 components.

Another significant trend to emerge at this stage was 'custom design'. This was the manufacture of an integrated circuit with just one customer and generally one equipment in view. The economics of this were certainly not fully understood initially and the balance sheets of some companies suffered until the lessons had been learnt. The major lesson was that most equipment manufacturers overestimate the requirement for their product and underestimate the time needed for development.

4 Enter M.O.S. (1966)

After its false start in 1964, m.o.s. began again to make its presence felt in 1966. The advertising claimed that the m.o.s. device offered the same advantages over the double-diffused integrated circuit that the transistor had held over the vacuum tube. These were: improved reliability, reduced component cost, savings in system size, weight and power, and potentially savings in overall system cost. In the writer's opinion it is true to say that, with the possible exception of the first, all these claims have been met.

The advantages as seen in 1966 are shown in Tables 1 and 2. M.o.s., however, did have one disadvantage over bipolar and that was speed. This was not seen as being a major disadvantage since over 90% of all logic systems at that time were operating below 1 MHz clock rate (i.e. within the capabilities of m.o.s.).

The big disadvantage that m.o.s. had in 1966 was its premature release in 1964. The unreliability of the early circuits had soured people's attitude to m.o.s. considerably. Even though by 1966 most of the early problems of stability had been cured, difficulties still remained with protecting the circuits from external static fields. Gate protection diodes were in use but their performance was inadequate. This lack of confidence in m.o.s. caused its general acceptance to be slow and caused the downfall of some suppliers.

Table 1. Comparison of m.o.s. and double-diffused epitaxial fabrication processes

	M.o.s.	Double diffused
no. of diffusions	1	4+epi
no. of process steps	38	130
high-temperature process steps	2	10

Table 2. Comparison of adder-subtracter integrated with r.t.l. and with m.o.s.

	r.t.l.	m.o.s.
no. of packages	18	1
silicon surface area (square mils)†‡	25 800	3200
size and weight reduction with m.o.s.	—	18 to 1
cost reduction with m.o.s.		
present	—	20%
potential	—	90%
power consumption (mW)	180	1 to 80‡

† 1 sq. mil = 645 μm^2. ‡ Power is a function of clock rate.

4.1 *Technology*

All initial m.o.s. production processes were of the p-channel enhancement variety and a cross-section of a typical process is shown in Fig. 8. The technology involved was basically that of the planar transistor with the addition of a gate oxidation stage. The m.o.s. transistor being a surface, as opposed to bulk, operated device, did call for attention to be concentrated on different factors. The stability of the oxides formed during the processing was a major factor.

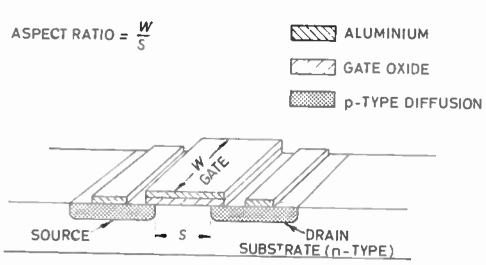


Fig. 8. The m.o.s. metal gate process. The source and drain consist of p-type regions diffused into a high resistivity n-type substrate. A thin oxide is grown over the region between the source and drain so as to just overlap these regions, and the top of this oxide is then aluminized to form the gate.

It was these effects that caused the 1964 difficulties. Each company had its own individual approach to solving the problem. The solutions used numbered amongst them, phosphorus stabilization, 'clean' oxides, and oxide-nitride composite gates.

The threshold voltage of these initial processes was in the region of 4–5 V and, since for good circuit operation power supply voltages needed to be some 4–6 times the threshold voltages, it was thought desirable to reduce the threshold voltage. Two methods were chosen for this (i) the oxide-nitride composite-gate structure mentioned above which reduced threshold voltages by changing the 'work function' potential between the metal and the gate dielectric, (ii) changing from (111) to (100) orientation for the silicon substrate which lowered the 'surface state density' of oxide to silicon interface and thereby reduced the threshold voltage. Both these changes were used to provide production processes with threshold voltages in the region of 2 V.

4.2 *Circuits*

As with previous integrated circuit technologies, initial emphasis was on demonstrating its widespread capabilities. As a device, the big advantages of an m.o.s. transistor were its very high input impedance and its zero source-drain offset voltage. It therefore rapidly found application as an analogue switch. As the complexity possible on the process increased, multi-channel analogue multiplexing circuits appeared. A few manufacturers constructed amplifier circuits using m.o.s. transistors but without commercial success, and it was soon realized that the major impact of m.o.s. technology would be in digital circuits.

The first important realization was that the m.o.s. transistor could form its own load device, and this, together with the way that all the device characteristics of the m.o.s. transistor scaled with surface geometry, made design very easy: the gain of an inverter circuit was the square root of the ratio of the gate widths. Because of the zero offset characteristic of the m.o.s. transistor, it was also possible to effect series as well as parallel gating and occasionally both within the same inverter circuit. This enabled considerable area reductions to be made. As a general figure, area reductions over double diffused technology of 10 to 1 were easily accommodated.

The gate of the m.o.s. device is electrically isolated from its source and drain giving extremely high input resistance ($\sim 10^{14}$ ohms) and since the gate impedance is essentially capacitive, the gate capacitance can be used to store voltage. This facility was used two ways, first to retain charge on the gate of the inverter transistor, and, secondly to reduce power by switching off the load transistor. An example of this, a two-phase shift register bit, is shown in Fig. 9. This stored charge would not remain indefinitely and minimum clocking frequencies were required. The typical operating range for the circuit shown would be 10 kHz to 1 MHz. This new development was called dynamic logic.

The amount of logic possible on an m.o.s. integrated circuit increased rapidly until by mid-1966[8] 100-bit shift registers were being made using dynamic logic, and dual decade counters were being made using static logic.

The challenge to reduce the size of logic circuits and the power they consumed continued and around 1967 the four-phase m.o.s. circuit was invented. Although these circuits used more transistors per bit, each transistor was of minimum size and power was consumed only during clock transitions. Table 3 compares integrated circuit approaches in 1968 from the viewpoint of an m.o.s. designer.

5 L.S.I. and All That

With the general acceptance, around 1968, that integrated circuits could compete on a cost basis with any other form of equipment construction method, the industry turned its attention to publicizing the scale of integration used.

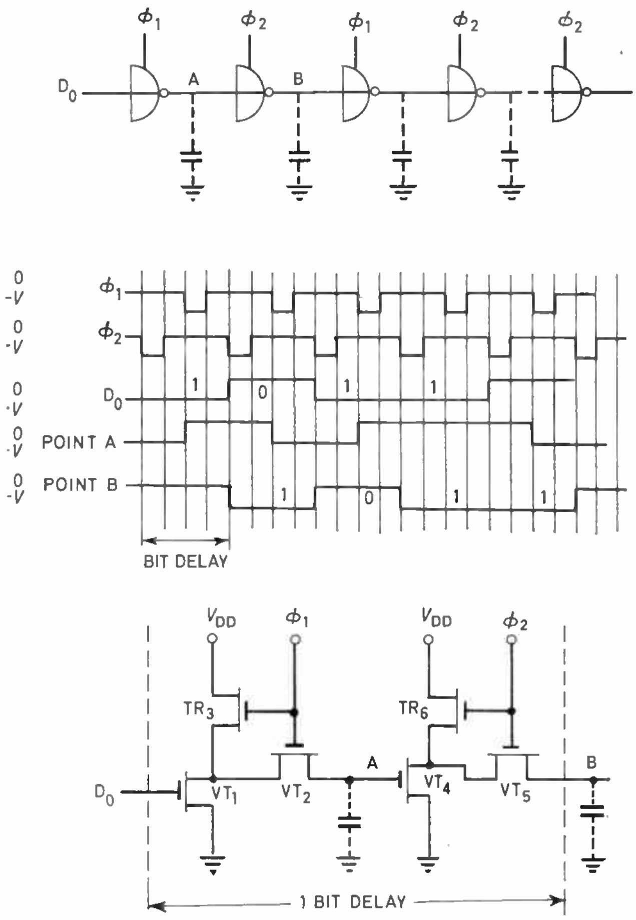


Fig. 9. A dynamic two-phase m.o.s. shift register.

What had previously been a steady progression, with the chip size of the integrated circuit increasing, and the size of the individual components reducing, was suddenly presented as a step function. Large scale integration (l.s.i.) had been invented!

Because of the greater complexity available in m.o.s., this technology became the front runner. To find outlets for the enormous complexity of logic that was now possible became a matter of great concern to the industry. Long serial-shift-registers were the first l.s.i. elements, mainly because of their very ordered design and the small number of input/output connexions required. By 1969 2000-bit shift registers had been announced containing around 12 000 individual m.o.s. transistors.

In tackling most system problems, however, it was generally found that the major problem was in partitioning the system so that the number of leads available in the package were not exceeded. This problem led to the development of 40 lead packages (200 leads were suggested). These packages were, of course, expensive.

An essential ingredient to l.s.i. was thought, at this stage, to be multi-layer metallization, some manufacturers producing circuits with three and occasionally four layers of metal. However, production problems have been considerable, and to this date the number of circuits available with multi-layer is very small. Of those available almost all are on bipolar technology. This does not mean that the requirement for multi-layer is not there, in fact the size of most circuits is controlled by the area taken by the metallization.

The big successes of l.s.i. are undoubtedly the random-access-memory (r.a.m.) and the calculator. Within a very few years, the number of i.c.s required to make a calculator has fallen from about six to one, and now the complexity of the calculator is increasing considerably within the single chip. Figure 10 shows the chip photograph of a single chip calculator. Semiconductor r.a.m. technology has now reached the state where, when compared at the system level, costs are comparable to, if not below, those of competing technologies. Four-kilobit memory chips are now becoming available.

6 Today

An outsider could be excused for thinking that after some fifteen years of integrated circuits, a general pattern to the industry should be emerging. This is not true, however; during recent years some of the basic principles on which the industry has been based have been questioned, namely, that m.o.s. is slow and cheap, and that bipolar is fast and expensive. Recent claims show that m.o.s. can be fast, and that bipolar circuits can be complex and cheap. The speed claims for m.o.s. come mainly from the recent work on n-channel[9] and complementary m.o.s. circuits,[10] and the complexity claims for bipolar circuits mainly from articles by B. T. Murphy[11, 12] on cheap bipolar processes.

6.1 *Technology*

6.1.1 M.o.s.

Production processes used today include the high and low threshold metal gate processes previously mentioned with the more recent addition of the silicon gate process.[13]

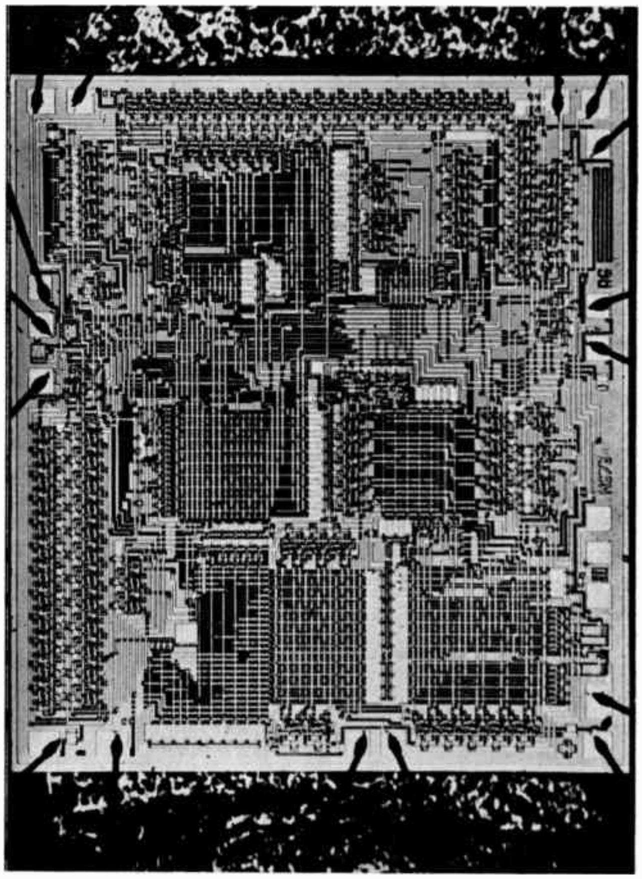


Fig. 10. A single chip calculator in m.o.s. (1971). (Chip size 3·6 mm × 4·1 mm).

Table 3. Comparison of various integrated circuit approaches (1967)
(Figures given to within a factor 2)

Approach	minimum speed	maximum speed	2-input NOR gate area (μm²)	bistable element area (μm²)	relative† size of logic system	mean power dissipation per logic gate max. speed	mean power dissipation per logic gate 10 kHz	noise immunity
bipolar (various)	d.c.	10–100 MHz	~20 000	~50 000	2–4	1–20 mW	1–20 mW	¼–2 V
d.c. m.o.s. i.c.	d.c.	500 kHz	~8000	~20 000	1	~2 mW	~2 mW	1–3 V
two-phase dynamic m.o.s. i.c.	10 kHz‡	1 MHz	~8000	~12 000	½	~2 mW	~20 mW	1–3 V
four-phase dynamic m.o.s. i.c.	10 kHz‡	3 MHz	~6000	~8000	⅓	~300 μW	~1 μW	1–3 V

† Relative sizes of logic system take into account the savings in logic possible with synchronous systems.
‡ The minimum frequency refers to clock frequency rather than output frequency, which can be d.c.

In the silicon gate process, the metal gate of the m.o.s. transistor is replaced by a layer of polycrystalline silicon. This process has the following advantages:

(1) The polycrystalline silicon can be used as a conductor giving something of the capabilities of two-layer metallization, but without the cost.

(2) Low threshold voltage.

(3) Auto-registration. With this the m.o.s. transistors are fabricated after the gate area has been defined giving greatly reduced inter-electrode capacitances and better process control.

These advantages lead to silicon gate circuits consuming some 40% less area than that required for other low threshold processes. The penalty for this improvement is the increased complexity of the process, at least one extra photoengraving stage being required.

Another process now arriving in production, in this case after almost ten years in development, is complementary m.o.s. In this process both n and p channel enhancement devices are fabricated together. This process is considerably more complex than traditional m.o.s. and circuits consume more area; however, the speed obtainable is close to that of current bipolar t.t.l. circuits and it remains to be seen whether or not its processing costs exceed those of the t.t.l. bipolar processes.

The work on improving the reliability of m.o.s. devices with gate dielectrics consisting of silicon nitride and silicon dioxide led to the development of the m.n.o.s. transistor.[14, 15] This transistor has a gate dielectric consisting of a thin (~ 20 Å) silicon dioxide region close to the silicon surface with a thick (~ 800 Å) silicon nitride overlayer. The interface between the two dielectric layers is capable of storing charge with a very long decay time. (Thousands of years have been predicted.) By inserting and removing charge from this interface, the threshold voltage of the m.o.s. transistor can be changed by electrical means. The m.o.s. device thus formed has non-volatile storage properties (i.e. it retains its preset state after power supplies to the device have been removed) and can be used as a single transistor storage element.

6.1.2 Bipolar

The bipolar digital scene at this moment is dominated by t.t.l. The only other family of logic still remaining is e.c.l., mainly because it offers gate delays lower than those possible with t.t.l. The fact that the circuit technique has not changed does not mean that the technology has remained static. A major technological change in t.t.l. has been the introduction of the Schottky diode.[16] This has enabled some manufacturers to remove a major yield factor from their process, namely gold doping. The Schottky diodes, when applied between the collector and base junctions of a transistor, prevent the transistor from saturating; this allows 'fast' and 'slow' transistors to be used within the same circuit. The Schottky diode is formed by making an unsintered connexion between the metallization, usually aluminium, and the low concentration n-type epitaxial region.

A radically new process capable of producing t.t.l. is that suggested by Murphy,[11,12] collector diffused isolation (c.d.i.). The process produces circuits of smaller area and with a lower process cost than conventional t.t.l. processes. Figure 11(a) shows a cross-section of a c.d.i. transistor.

The recently developed techniques of selective etching and oxidation have produced a variety of new processes in recent months,[17, 18, 19] the most significant at this stage being the 'isoplanar' process of Fairchild; a cross-section of this is shown in Fig. 11(b). This process claims m.o.s. packing density with comparable process costs, but with significant speed advantages.

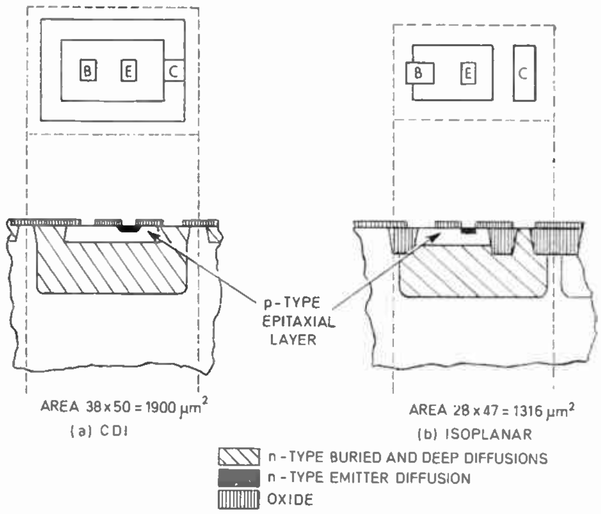


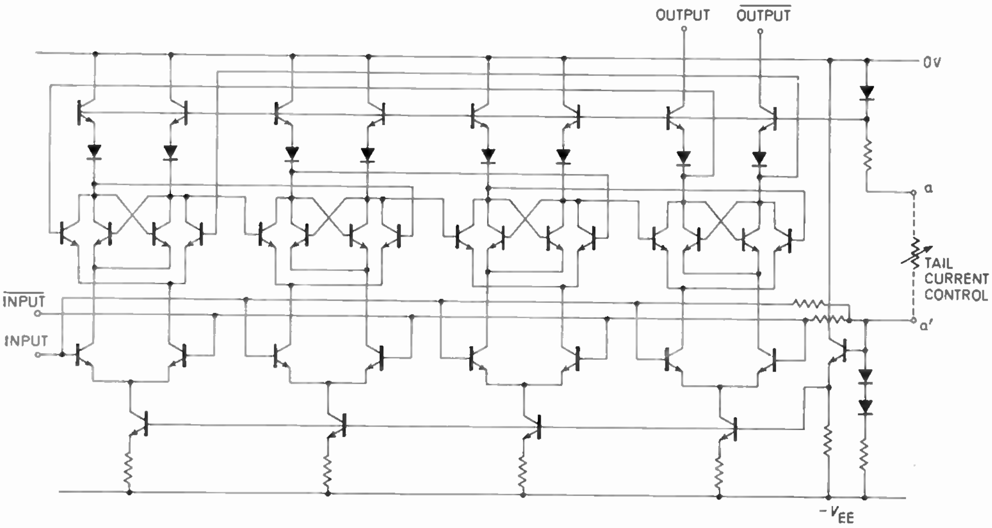
Fig. 11. Simplified bipolar technologies. Plan view and cross-sections of c.d.i. and isoplanar processes are shown together with the area of isolated transistors made on these technologies.

6.2 *Circuits*

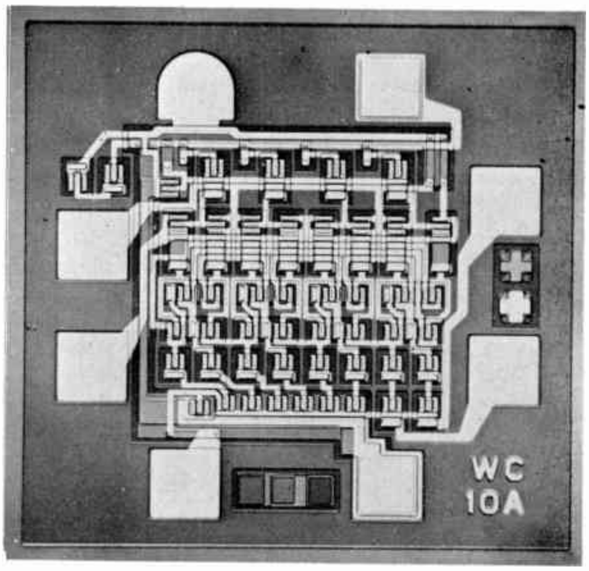
6.2.1 Digital

To a large degree the dominance of t.t.l. has been attributable to aggressive marketing and the provision of the necessary circuit and applications information to permit the systems designer to produce his system. The range of t.t.l. devices now available exceeds 100. The pricing policies of the t.t.l. manufacturers have been such that only the very efficient could remain in the business—this, however, has been of great benefit to the user. The major developments outside this field have been in m.o.s. where l.s.i. has been the vogue, and in e.c.l. where higher speed has been the keyword.[20] Whilst clock rates of 75 MHz are possible with t.t.l. circuitry, divider circuits with clock rates of 1000 MHz are now possible using e.c.l. techniques. Figure 12 shows a diagram and chip photograph of such a circuit. The major impetus for this work is the movement towards higher data rates in pulse-code modulation (p.c.m.) systems.

An important stimulus for cheap bipolar processes has come from the possibility of converting the vast memory systems market to semiconductors. One-kilobit bipolar memory circuits are now available at speeds in excess of those possible with magnetic cores or m.o.s.



(a) Divide by four circuit.



(b) Chip photograph of a 1 GHz counter.

Fig. 12. The circuit diagram and chip photograph of a digital divider capable of input frequencies in excess of 1 GHz (1972). It is of interest to note that the circuit contains no resistors in the active portion. The circuit divides by four giving complementary current outputs. Chip size $0{\cdot}75 \times 0{\cdot}75$ mm).

In the m.o.s. field the new non-volatile memory device described in Sect. 6.1.1 is very promising for use in memory systems. Predictions of up to 16 kilobits per chip have been made; however, the technology has yet to reach production status.

6.2.2 Analogue

From a slow beginning, the analogue device is now increasing its sector of the total integrated circuit scene. Circuits are now finding application in such consumer items as washing machines, cars and television sets. Radio-communication systems are now almost all i.c. based and a new range of higher frequency integrated circuits are just coming onto the market.[21] These will certainly increase the penetration of analogue circuit function.

7 Trends for the Future

7.1 *Speed*

One certainty for the future will be the requirement for faster logic. If the impetus for this does not come from the computer manufacturer, it will come from the necessity for more data transmission capacity in telecommunications. One caution here is that because of the length of connexions needed in the assembling of packages, there may be a requirement to increase the complexity of the function to beyond that dictated by maximum chip handling power. This may require compromises to be made to obtain the best system speed. Gate speeds of 0·5 ns and clock rates of 1 GHz should be possible within the next few years.

7.2 *Complexity*

Already it is possible to produce all the functions for simple calculators on one i.c. chip, and it is a simple prediction to say that even the most complex of calculators will be single-chip within a year or so. Having achieved this, the next development would appear to be the single-chip mini-computer. Memory circuits will undoubtedly progress so much in complexity that anything one might predict now will look foolish within a short space of time. One development that may come from work being carried out in the memory field is the l.s.i. general-purpose logic block. Using m.n.o.s. devices, a reprogrammable logic block, based on the programmed logic arrays now being made, could be made to vary its function under program control to produce a self-ordered system.

Complex functions combining both digital and analogue functions could also become common.

7.3 *Bandwidth*

Fairly comfortable predictions can be made that amplifiers with bandwidths up to 1 GHz should be possible and with gain-bandwidth products in excess of 10 GHz.

7.4 *New Techniques*

Two new techniques, of which more may be heard in the future, are injection logic,[22, 23] and charge-coupled-devices (c.c.d.).[24, 25] Injection logic appears to offer a viable alternative to m.o.s. for low speed, high complexity, low cost applications, and it may be possible to improve the speed. Predictions of efficiency (i.e. power delay product) show considerable benefits over m.o.s. C.c.d. is one of a number of new approaches for delaying analogue signals in a digital manner. The basic c.c.d. process requires no diffusions and operates by moving charge between potential wells in the silicon substrate. This technique shows considerable application to the area of optoelectronics, and could lead the way to the all-solid-state vidicon.

Both these techniques are very new and it would require a brave man to expose their limitations at this time. One need only recall the many who 'wrote off' the transistor in the early 1950s and the integrated circuit in the 1960s.

8 Acknowledgment

The author wishes to thank the Plessey Company Limited for permission to publish this paper and to pay tribute to his colleagues throughout the industry, without whom the miracle of integrated circuits would not have been possible.

9 References

1. Holt, W., 'A solid circuit amplifier', Paper presented at National Conference on Solid Circuits and Micro-miniaturization (NCSCM) at West Ham College of Technology, June 1963.
2. Dickens, H. K., 'Parasitics in integrated circuits', *Electronics*, pp. 32–36, 5th June 1963.
3. Boulter, B. A., 'Solid circuit elements using multiple-emitter transistors', NCSCM, West Ham College of Technology, June 1963.
4. Thompson, P. M., (Plessey Co. Ltd.), Patent Application 24222/61.
5. Widlar, R. J., 'A Monolithic Operational Amplifier', Fairchild Applications Bulletin, APP105, 1964.
6. Gay, M. J. *et al.*, 'Integrated circuits for professional communications systems', ISSCC Digest of Technical Papers, pp. 10–11, February 1967.
7. Brothers, J. S. and Gay, M. J., 'Monolithic integrated circuit capacitors', *Proc. IEEE*, **53**, No. 6, p. 644, June 1965.
8. Bogert, H. Z., Jr., 'Metal oxide silicon integrated circuits', *SCP and Solid State Technology*, pp. 30–35, March 1966.
9. 'Special report on n-channel m.o.s.,' *Electronics*, 8th May 1972.
10. Boleky, E. J., 'Subnanosecond switching delays using c.m.o.s./s.o.s. silicon gate technology', ISSCC Digest of Technical Papers, p. 225, February 1971.
11. Murphy, B. T. and Glinski, V. J., 'Transistor–transistor logic with high packing density and optimum performance at high inverse gain', ISSCC Digest of Technical Papers, pp. 38–9, February 1968.
12. Murphy, B. T., *et al.*, 'New simplified bipolar technology and its applications to systems', ISSCC Digest of Technical Papers, p. 150–1, February, 1969.
13. Faggin, F. and Klein, T., 'A faster generation of m.o.s. devices', *Electronics*, pp. 88–94, 29th September 1969.
14. Frohman-Betchkowsky, D. and Lenzluger, M., 'Charge transport and storage in metal nitride silicon (m.n.o.s.) structures', *J. Appl. Phys.*, **40**, pp. 3307–19, 1969.
15. Vieth, R. F., 'Nitride-oxide layer proofs memory against data loss', *Electronics*, 5th July 1971.
16. Kronlage, J. W., 'Schottky clamped t.t.l.', IEEE International Convention Digest, pp. 144–5, March 1971.
17. Peltzer, D. and Horndon, W., 'Isolation method shrinks bipolar cells for fast, dense memories', *Electronics*, p. 53, 1st March 1971.
18. Appels, J. A. and Paffen, M. M., 'Local oxidation of silicon: new technological aspects', *Philips Research Reports*, **26**, p. 157, June 1971.
19. Mudge, J. and Taft, K., 'V-ATE memory scores a new high in combining speed and bit density', *Electronics*, p. 65, 17th July 1972.
20. 'Focus on fast logic,' *Electronic Design*, **12**, pp. 50–7, 8th June 1972.
21. Blomley, P. F., 'A wideband analogue multiplier', to be presented at ISSCC, February 1973.
22. Berger, H. H. and Wiedman, S. K., 'Merged transistor logic—a low cost bipolar logic concept', ISSCC Digest of Technical Papers, pp. 90–1, February 1972.
23. Hart, C. M. and Slob, A., 'Integrated injection logic—a new approach to l.s.i.', ISSCC Digest of Technical Papers, pp. 92–93, February 1972.
24. Session XII, ISSCC Digest of Technical Papers, February 1972.
25. Altman, L., 'The new concept for memory and imaging: charge coupling', *Electronics*, 21st June 1971.

Manuscript received by the Institution on 2nd November 1972. (Paper No. 1494/CC.158.)

UDC 621.382

The struggle for power, frequency and bandwidth

CARL S. DEN BRINKER*

SUMMARY

The relationship between physical and electrical parameters is reviewed. Some of the newer technologies are discussed. The interdigitated structure is compared with the two main alternatives offered at present. At the same time a basic correction is made to the well established approximation of power gain as applied to power devices. The necessity for this correction is also developed with respect to modules, leading to the conclusion that thin film modules offer probably the only chance to utilize fully the intrinsic power gain capability of devices. As the useful limit of silicon appears now in sight, numeric examples are used wherever possible.

*Texas Instruments Ltd., Manton Lane, Bedford.

1 Introduction

An insatiable hunger for bandwidth and power is symptomatic of an advancing technology. The ability to convey more and more intelligence of one sort or another cannot be separated from technological progress. It is therefore not at all surprising that the transistor technology has, since its inception in 1948, moved continuously towards higher frequencies and higher power levels.

What does not appear quite as self-evident, is that semiconductor technology has also managed to repeat what appeared possible before but with more effectiveness. In this review some of the aspects which have brought this change about will be reviewed. It is only by a thorough appreciation of the detailed changes that are taking place, that one can fully utilize the potential of the latest technologies.

Compare, for example, the first c.w. transmitters employing spark gaps, with a modern frequency synthesizer controlled communication link. Both systems tend to have a considerable bandwidth capability. During the period that existed between these two types of systems being in existence, an effort was made to reduce unwanted transmission by multi-stage tuning. A complete circle has been made from uncontrolled bandwidth to narrower band systems and now it appears that the controlled channel, broadband systems are flourishing. As the frequency channel controls tend to become increasingly more complex they must, of necessity, appear only once in the system. Therefore as multi-stage tuning becomes prohibitive in cost alone, the transmitter requirements, in terms of linearity of phase etc., become increasingly more stringent.

Besides the question of cost there is another major factor to consider: speed of operation. For example, in a military system it may be desirable to hop from channel to channel for reasons dictated by electronic counter measure techniques. Even the most capable human operators would find it difficult accurately to tune successive stages for 1 MHz channel spacing when the mid-carrier frequency is, say 4 GHz.

In this review on power, bandwidth and frequency it is intended to review first the physical characteristics of power devices. This will be followed by a brief review of the parametric behaviour of h.f. power transistors. The concluding part of the paper will be devoted to the application of these devices.

2 Fundamental Geometry Considerations

The transfer relationship between the applied base-emitter voltage and the resulting collector current[1] is very similar to the VI relationship of a simple junction diode, i.e.

$$I_c = \alpha T^{1\cdot 4} \exp\left[\frac{q}{kT}(V_{be}-\phi_0)\right] \qquad (1)$$

where $\alpha T^{1\cdot 4}$ has the dimensions of current α is a constant depending on the device fabrication and ϕ_0 is the silicon energy gap at 0 kelvins.

Two temperature terms appear in this expression. The first is in the multiplying coefficient, resulting in a collector current rise with temperature when V_{BE} is held constant. The second temperature term appears in the exponent of the exponential term leading to a reduction in current. Of these two, the first one dominates. As, in fact, the multiplying coefficient is directly related to n_i^2 this term leads to an increase of approximately 15% deg C^{-1} in silicon. This again results in a typical temperature coefficient of $-1{\cdot}75$ mV deg C^{-1} for the total transfer function described in equation (1). Having discussed the similarities between a simple diode and a transistor, we should now consider some of the fundamental differences. In Fig. 1 a simple transistor structure is shown.

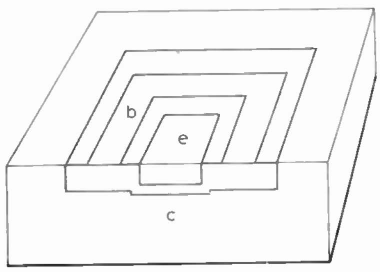


Fig. 1. Simple transistor structure.

Under steady-state conditions a given base charge will sustain a given collector current. This collector current can only be maintained as long as hole-electron pairs are replenished at the rate at which they recombine. This replenishment is achieved by the base current. As the base current is a majority carrier current, it is drift controlled, and it will be subject to an electrostatic profile. Close to the surface the contact potential ϕ, of the emitter-base junction will be considerably higher than at the bottom, due to the higher density of donor and acceptor atoms near the surface.[2] As a result, emission of carriers will occur mainly at the bottom end of the emitter surface. The bulk of the carriers emitted by the emitter are caught by the collector due to its proximity. However, this very essence of transistor action, i.e. the narrow base width, results in the emitter edge crowding effect.

The cause of crowding is most dramatically illustrated by a numeric example. After base diffusion using a normal process, one would typically measure a surface resistivity of 180 Ω/square and this is the resistivity that the base current has to overcome in the base-contact to emitter-edge path. In the active base area of Fig. 1 one expects a marked increase in the surface resistivity as measured at the bottom surface of the emitter. That is to say, if one could measure it directly! In fact, if the base region were homogeneous, one would expect an increase of the order of just under three times in resistivity as the base width is about a third of the depth of the total base diffusion. By indirect means one can, in fact, measure a surface resistivity of the order of 30 times greater, typically 5 kΩ/square. This is entirely due to the rapidly increasing resistivity as one traverses the base region from the device surface to the bottom surface of the collector junction.[2] With this disparity it is not possible to achieve a forward bias at the centre of the emitter surface.

Consider again a simplified version of equation (1):

$$I_c = I_0 \exp\left[\frac{qV_{be}}{kT}\right]. \tag{2}$$

To achieve an increase of current V_{BE} needs to be raised according to

$$xI_c = I_0 \exp\left[\frac{q(V+\Delta V)_{be}}{kT}\right]. \tag{3}$$

If x is solved by substitution of (3) into (2) one obtains the solution

$$\Delta V_{be} = \frac{kT}{q} \ln x. \tag{4}$$

The difference voltage for a current ratio of 2:1 can be resolved from (4) by putting x equal to 2 and solving kT/q. The latter is approximately 26 mV at room temperature. Thus:

$$\Delta V_{be}\Big|_{I_1/I_2=2} = 26 \text{ mV} \times 0{\cdot}6931 = 18 \text{ mV} \tag{5}$$

Thus, only a voltage drop of 18 mV is needed along the bottom surface of the emitter before the emitter emission per unit area is halved. Similarly, (4) can be solved for other current ratios.[3]

If only the edge of the emitter junction participates in transistor action, then the only path open to achieve high current capability is to increase the emitter periphery. Merely increasing the surface area decreases both the yield and the upper frequency of operation.

The oldest and best known solution is shown in Fig. 2, i.e. an inter-digitated structure. As early as 1958 optimum dimensions were analysed by J. W. Early.[4]

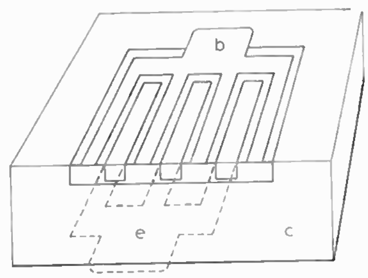


Fig. 2. Interdigitated structure.

It may at first sight be necessary to achieve the largest possible ratio of emitter periphery P_e to base area A_b, but this is not the only consideration. The first natural limitation to this philosophy is that long thin metallized stripes of the base and the emitter contacts tend to have a finite resistance. This resistance, if it is too high, would stop any transistor action at the end of the fingers. As shown before, only an 18 mV drop would be required to halve the junction current density. To reduce the diode current density to 10% at room temperature, only a 60 mV drop is required. This in fact shows how relevant the metallization is.

The last major consideration is due to the different temperatures that prevail along the width of the device bar. Heat flow from the centre of bar surface is more difficult than from the outer edges of the device. Therefore, the area in which the centre stripes are located will tend to heat up more. As a result of the relationship shown in equation (1), the centre emitter stripes will carry a higher current than the outer ones. This increases the heat again. Because of this cumulative effect, one may well find that the device dissipation is entirely limited by the central part of the device. Typically, if a 10 deg C temperature difference is maintained from the centre to the edge of the device, the result of equation (4) combined with the diode temperature coefficient of approximately $-1{\cdot}75$ mV deg C^{-1} will result in a 2:1 current density difference between centre and edge.

A solution to this problem is to introduce ballast resistors in series with the emitter stripes. In practice this is simply achieved by spacing the emitter stripe metallization some small distance from the emitter contact spine. This is shown in the sectioned drawing of Fig. 3. Also shown are some typical dimensions that one may encounter in v.h.f. and u.h.f. power transistors.

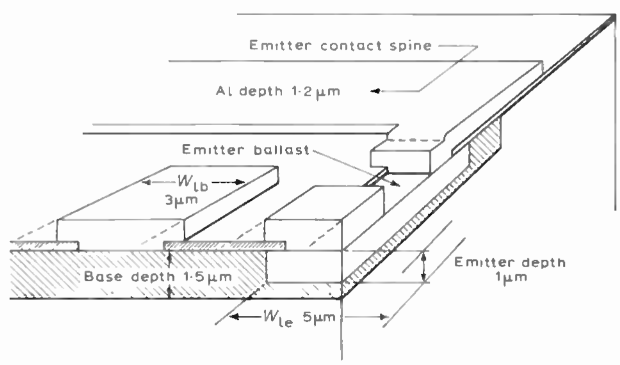


Fig. 3. Sectional view of power transistor with ballast resistors in series with emitter strips.

The ballast resistor also greatly improves the secondary breakdown protection and it is therefore common practice now to introduce emitter ballast resistors.

3 The Parametric Implication of Physical Dimensions

The geometric details so far discussed have been essentially associated with power. Consideration must now be given to the dynamic properties. For example, in the previous chapter the high donor and acceptor density near the device surface was given as the reason why carrier injection is unlikely to take place near the surface. Typically the doping concentration changes by some three decades from 10^{20} cm^{-3} at the device surface to 10^{17} cm^{-3} at the bottom surface of the emitter. As a result the capacitance of the emitter side walls is 1000 pF mm^{-2} whereas the bottom surface only produces 350 pF mm^{-2}. The collector surface capacitance is even lower still at typically 120 pF mm^{-2}.

The implication of these values is obvious. When long emitter stripes are used of say 3 μm width and 1·5 μm depth, 26% of the capacitance will be due to the bottom surface and 74% due to the side wall surfaces. The surface area can be reduced to improve the frequency response of the device, but by far the most effective way to improve the frequency response will be (a) to reduce the diffusion depth, and/or (b) to minimize the profile change. It is for the latter reason that so much emphasis is placed on ion implantation techniques.

The cut-off frequency of transistors is a function of a number of contributing factors and probably the simplest method to show this was introduced by Beaufoy and Sparkes.[5] This was subsequently followed up by the more generalized analysis of E. O. Johnson *et al.*[6] Basically Johnson states that any network having an angular cut-off frequency relates the time elapsed before an input signal arrives at the dissipation elements,

i.e.

$$f_t = (2\pi\tau_t)^{-1} \tag{6}$$

where τ_t is the mean transit time of a charge carrier, traversing the element at a mean velocity v.

If one considers again Fig. 1 it can be seen that four separate transition regions will influence the total transit time.

(a) The emitter base junction (τ_e)

The voltage–current relationship at the emitter junction is defined by the transconductance g_m while the charge alignment is controlled by the transition capacitance C_{te}. This yields an emitter transit time

$$\tau_e = C_{te} g_m^{-1}$$

where

$$g_m^{-1} = \frac{kT}{qI_e}. \tag{7}$$

(b) Base transport (τ_b)

To sustain a given collector current, a given base charge is required as mentioned earlier,

$$\tau_b = Q_b/I_c = \frac{W^2}{nD} \tag{8}$$

where W_b = base width, D = diffusion constant of the carriers in the base, and n is a constant, whose values can lie typically between 2 and 5. Several approximations to n are possible.[7] But as we are considering very high current densities n will approach 2.

(c) Collector junction transport (τ_c)

If one assumes that the collector junction is only lightly doped and that a reverse bias is applied to this junction, one cannot ignore the junction width and the transit time through the depletion layer.

$$\tau_c = \frac{W_{tc}}{2v} \tag{9}$$

where W_{tc} = the collector transition region width and v is the saturated velocity of carriers.

(e) Collector bulk effect (τ_B)

This last factor arises from the collector base capacitance C_c and the bulk resistance of the transistor bar from the collector junction to the bottom of the device bar, i.e.

$$\tau_B = C_c r_c. \tag{10}$$

In order to minimize the latter effect r_c and C_c must be minimized. To minimize C_c a low impurity density must prevail, i.e. the collector body material then has a high resistivity, and this defeats the object of a low r_c unless an epitaxial structure is used. What is also relevant is that high resistivity material allows much larger voltage peaks to occur across the collector junction. As large spikes are likely to occur when mismatches occur, it follows that a fairly thick epitaxial layer will be required to cater for high v.s.w.r. conditions.

When summing the total transit time conditions it is very worthwhile to consider how much the separate factors contribute to the total for a typical, 'well designed' device, e.g.

$$\begin{aligned}\tau_t &= \tau_e + \tau_b + \tau_c + \tau_B \\ 100\% &= 40\% + 10\% + 45\% + 5\%.\end{aligned} \qquad (11)$$

(This applies to the small-signal device of Fig. 1.)

From the foregoing discussion it appears that increasing the voltage capability does not enhance the transit time. In fact, Johnson pointed to this fact as early as 1959 and one is reminded to the limit condition in the review paper by Gri.[8] The Johnson limit that prevails under idealized circumstances is

$$V_{max.} f_t = \varepsilon\, v_s / 2\pi \qquad (12)$$

where ε is the dielectric breakdown field and v_s is the saturated drift velocity.

Values of $V_{max} f_t$ for Si and GaAs are 2×10^{11} V s^{-1} and 1×10^{12} V s^{-1} respectively. Already a number of silicon devices show values greater than 1×10^{11} V s^{-1}.

So far we have neglected to look at the control terminal, except when we considered emitter crowding. It can readily be seen from Fig. 2 that for a given emitter finger length the total extrinsic base resistance r_b, is dependent on the distance of the base to emitter fingers and is inversely proportional to the number of finger pairs within the base area. Again improving the P_e/A_b ratio appears to be a dominant consideration.

4 The Effect of Packaging

One could envisage a number of solutions to the basic problem of connecting broad, low inductance leads to the relatively small device bars. Probably the most successful commercial package at present used is the JO package, the principle of which is shown in Fig. 4.

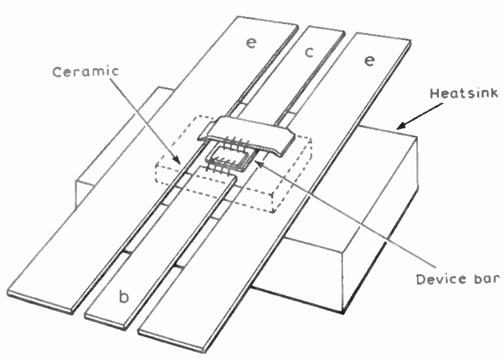


Fig. 4. JO package showing broad, low inductance leads.

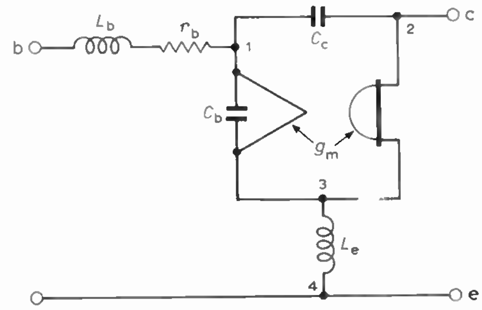


Fig. 5. Simplified circuit of u.h.f. power transistor.

The heatsink has a ceramic base fused to it. The normal material is beryllium oxide (BeO) as this material gives the highest thermal conductance. The collector terminal is mounted directly to the collector film on the BeO wafer, this results in minimizing the collector inductance. The emitter termination consists of two parallel strips connected via a bonding bar, which also bridges over the collector lead. This is probably the main improvement over the more conventional TO 117 type headers, as the earth path from base to emitter as well as from collector to emitter is now minimized. However it can readily be seen that no matter how wide the lead patterns are made the limitation with present day technologies is still dictated by the inductance of the bonding wires. A rather simple and elegant improvement will be discussed in Section 7.

5 The Electrical Performance of Power Devices

In spite of the difference between linear small-signal behaviour and the non-linear large signal behaviour one can evaluate the practical behaviour of large-signal conditions to a fair degree with linear analysis. As a starting point, a typical 1 GHz example will be used to show the rationale of some of the better known approximations.

The simplified circuit to be used is shown in Fig. 5, it consists essentially of a primitive h.f. hybrid π to which extrinsic base and emitter inductances are added.

Now the topology of the network is such that one expects the h parameters to come out as 'natural', i.e. input impedance and output admittance are the least frequency sensitive representations. Therefore one can insert or omit the $L_b + r_b$ branch whenever necessary. The main floating y matrix, excluding the base branch, becomes as shown in Table 1

Table 1

$j\omega(C_b+C_c)$	$-j\omega C_c$	$-j\omega C_b$	0
$-j\omega C_c + g_m$	$j\omega C_c$	$-g_m$	0
$-j\omega C_b - g_m$	0	$j\omega C_b + \frac{1}{j\omega L_e} + g_m$	$-\frac{1}{j\omega L_e}$
0	0	$-\frac{1}{j\omega L_e}$	$\frac{1}{j\omega L_e}$

Now the input admittance y_{11} with the output short-circuited can easily be derived as follows. Terminal 4 is earthed and thus row and column 4 are therefore eliminated. As the output is shorted to earth, column 2 and row 2 also disappear and now $y_{11} = \Delta/\Delta_{11}$ where Δ_{11} is the remainder (y_{33}) after row and column 1 have been eliminated, i.e.

$$y_{11} = \frac{j\omega(C_b+C_c)\left(j\omega C_b+\frac{1}{j\omega L_e}+g_m\right) - (g_m+j\omega C_b)(j\omega C_b)}{j\omega C_b+\frac{1}{j\omega L_e}+g_m} \quad (13)$$

The first term of the numerator in Table 1 is dominant, with the second significant term[3] being only of the order of 10% of the first (see Table 2).

Table 2

Typical 1 GHz device data

(BLW 32	10 W at 1 GHz)	
g_m	3·5 A/V	100 mA
C_b	350 pF	,,
C_c	4 pF	28 V
r_b	0·9 Ω	
L_e	0·4 nH	
L_b	1·2 nH	

Inverting Table 1 and adding $r_b+j\omega L_b$ to the reciprocal now yields

$$h_{11} = r_b+j\omega L_b+j\omega L_e+\frac{1}{j\omega C_b}+\frac{g_m}{C_b}L_e$$

$$\simeq r_b+\frac{g_m}{C_b}\cdot L_e+j\omega(L_b+L_e). \quad (14)$$

By substituting ω_t for g_m/C_b one obtains the well-known expression

$$\mathrm{Re}\,(h_{11}) = r_b+\omega_t L_e. \quad (15)$$

Before considering h_{22} a number of simplifying techniques can be applied to the determinant. First row and column 4 are deleted as terminal 4 is again identified. Inspection will show that considerable simplification can be achieved by the rule which states that 'the value of a determinant remains unchanged if to one row/column is added any linear combination of its other rows or columns'.

The expression for h_{22} can be derived by leaving the base branch open, i.e. the input is open circuit

$$h_{22} = \frac{C_c}{L_b}g_m(1+\omega^2 L_c C_c)+j\omega C_c\left(1-\frac{g_m^2 C_c L_e}{C_b^2}\right). \quad (16)$$

Again by judicious manipulation this can be reduced to

$$h_{22} = C_c\omega_t+j\omega C_c. \quad (17)$$

This latest expression is also well established. It is worthwhile now to assume that the $r_b C_c$ product is minimized for maximum frequency. Numerically from Fig. 6, $r_b . C_c$ is $0{\cdot}9\times4{\cdot}5\times10^{-12} = 4$ p s. This is much less than the inverse radial frequency of operation of 0·6 n s ($1/2\pi$ GHz). As a result one can therefore represent the transistor of Figs. 5 and Table 2 by the approximation of Fig. 6.

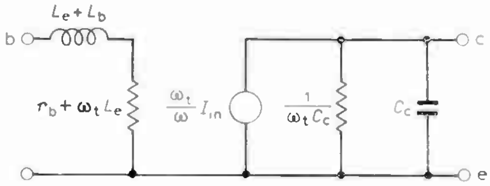


Fig. 6. Approximate equivalent of Fig. 5 for maximum frequency operation.

This circuit can now be used to evaluate the power gain under conjugate matched conditions.

The generalized expression for transducer gain is

$$G(f) = \frac{P_0}{P_i} \simeq \frac{I_0^2(1/\mathrm{Re}\,(y_{out}))}{4I_i^2(1/\mathrm{Re}\,(y_{in}))}. \quad (18)$$

From the previously derived approximate solution this can now be rewritten as:

$$G(f) \simeq \frac{|h_{fe}|^2}{4}\cdot\frac{1}{\omega_t C_c(\omega_t L_e+r_b)} \quad (19)$$

Substitution of the parameters by the values given in Table 2 yields a gain of 3·5 or 5·5 dB. This is not excessively different from the actual gain of 6 dB.

However, if the effect of emitter inductance could have been eliminated using the better known approximation of gain,[10] the result would have been approximately 11 dB. These figures speak for themselves.

6 Improvements in Devices

In the earlier chapters a number of factors were discussed, which would improve power gain, cut-off frequency and the power handling capability of devices. One of the widest discussed subjects is the emitter geometry. Basically three types of geometry are at present used:

(a) Overlay, which employs broad emitter metallization for a minimal emitter area (see Fig. 7).

(b) Interdigitated structures, which have been illustrated in Fig. 2. They are probably the commonest in practical use.

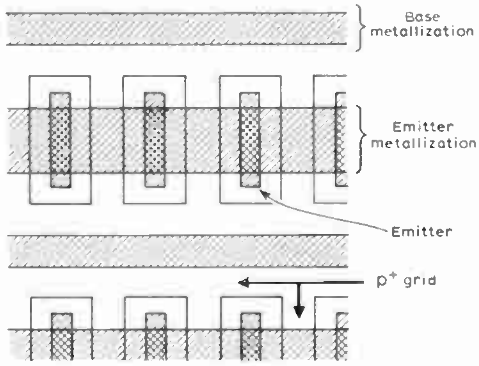


Fig. 7. Overlay technique for broad emitter metallization.

(c) Matrix grid structures, which have the advantages of broad metallization and high P_e/A_b (see Fig. 8).

The overlay structure is the least effective in achieving high P_e/A_b aspect ratios. The limit value is typically 3·5.[11] However, this technique will improve the problem of potential drop along long emitter fingers. The interdigitated structures have a typical aspect ratio limit of 6. At higher frequencies the problem of potential drop along the emitter fingers diminishes. For example, if at frequencies of $f_t/2$, the emitter width is approximately three times the base stripe width, transverse equilibrium will prevail as $l_e \simeq 3l_b$. For this reason, a larger part of the active base area is used at higher rather than lower frequencies, resulting generally in higher power handling capability at frequencies near f_t than at lower frequencies. Using a basic interdigitated structure, small-signal devices have been reported[12] having an f_t of 6 GHz and a noise figure of 3·6 dB at 4 GHz. This device has an emitter dimension of 1 μm.

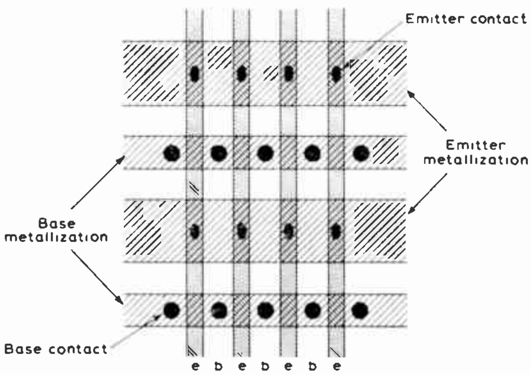


Fig. 8. Matrix grid structure.

The third major geometric structure provides probably the highest emitter periphery to base area aspect ratio at approximately 8 (see Fig. 8). The principle of the matrix grid structure is almost self-evident from this drawing. Very narrow emitter diffusion and broader base width stripes are connected by transverse metallization stripes. Production results of 5 W power output and a power gain of 4 dB at 4 GHz were reported in March 1972[13] using the matrix geometry. For all these improvements a price must be paid. First of all, the fine geometries require special mask techniques. The conventional E-line lens system of step and repeat imaging is now superseded by the G-line lens system. Non-reflective chromium working plates are needed to obtain micron geometry contact prints.

The processing of these fine structured devices also brings with it some problems. The first problem to be overcome is emitter run-on, i.e. the emitter diffusion pushing the collector further in, as is shown in Fig. 1. Arsenic emitter diffusion seems to overcome this and, for this reason, the As process is increasingly used for shallow diffusions.

The second series of problems is associated with the use of aluminium. Aluminium is the classical metallization metal because it is easy to use and it adheres well to both Si and SiO_2. However, it also reacts with these materials. With the shallow diffusions, sideways diffusion is minimized. Removing the glaze and phosphorus dopant source is a critical process and the original SiO_2 is precariously minimized over the emitter base junction. Subsequent Al metallization could further breakdown the SiO_2 with the subsequent shorting of the emitter to base junction (see Fig. 9). This is generally described under the heading 'emitter wash-out' and its effect can be reduced by depositing additional SiO_2 after diffusion, and before metallization, to protect the vulnerable region. A second alternative is to introduce a thin layer of Si_3N_4 between two SiO_2 layers, this produces a nitride shelf.

A second problem associated with Al is ion migration. Thermally activated metal ions will move, if they are subjected to an electric field, towards the cathodic end. This process can be slowed down by adding small amounts of Si (0·3%) or Cu (4%) to the aluminium or glass passivation of the Al.

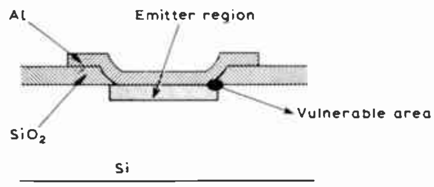


Fig. 9. Illustrating possible area of breakdown using aluminium metallization.

Probably the simplest expression which relates the reliability of devices to ion migration is due to Black:[14]

$$\text{m.t.t.f.} = \left(\frac{W \cdot t}{CJ^2}\right) e^{\phi/kT} \qquad (20)$$

where W = stripe width
t = stripe thickness
ϕ = activation energy for diffusion of the metal
C = constant
J = current density.

At present the panacea for the ills that aluminium metallization produces is thought to be gold. To avoid AuSi eutectic formation (this occurs at 367°C) a composite system is required employing barrier metals. The barrier metal should provide a good contact with Si, adhere well to SiO_2 and be a good barrier to Au. A number of these systems have been described in an issue of the *Microwave Journal* devoted to reliability.[15] It is claimed that the gold system gives an improvement of m.t.t.f. of some fifteen times over aluminium.

It may appear again that the sole preoccupation is with the emitter problems. This however is not true. If emitter ballasting is used the collector area can then be broken up into a collector cell structure. Without ballasting the phase differences of signals in different cells would result in different power drives. The division of the collector into cells tends to avoid hot spots and because of this, the multi-collector cell structure is almost common practice now in high-power microwave transistors.

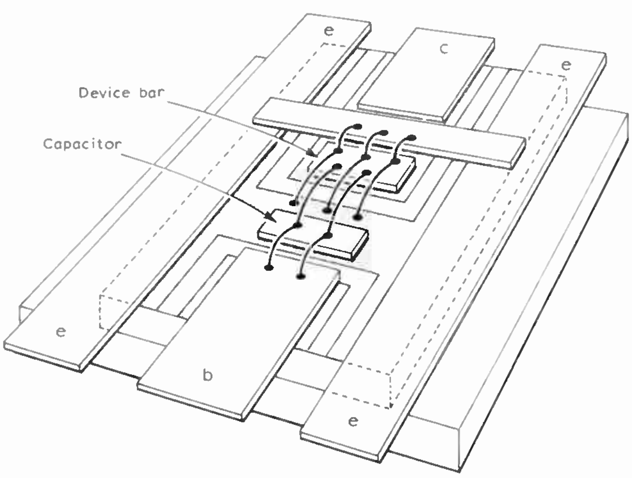
7 Improvements in Device Packaging

It was suggested earlier that the bonding wires can in the limit dominate the extrinsic inductance of power h.f.

transistors. In the practical 1 GHz case it was shown that the power gain was lowered by some 6 dB due to emitter lead inductance. Thus merely improving the device bar is not going to improve matters unless this is coupled with a drastic packaging improvement. An example of an improved JO construction is shown in Fig. 10 (T.I. experimental device). The emitter bonds are not only made to the bridging bar, a parallel set of emitter bonds is also made to the earth metallization on the ceramic substrate. The base bonds are first made to a parallel capacitor and then skipped to the base lead ribbon. The first improvement aids the power gain considerably, whereas the second modification improves the bandwidth of the device. From equation (14) it appears that the real and imaginary parts of h_{11} are likely to yield a Q in the order of 3 at frequencies close to f_t.

The relatively high Q that occurs in practice mainly limits the bandwidth. Again this is best illustrated by a simple example. From (14):

$$h_{11} \simeq r_b + \frac{g_m}{C_b} . L_e + j\omega(L_e + L_b)$$
$$= 0{\cdot}9 + 4 + j\omega(0{\cdot}4 + 1{\cdot}2) \times 10^{-9}.$$



At 1 GHz therefore

$$h_{11} = (4{\cdot}9 + j10)\,\Omega \quad \text{i.e.} \quad Q \simeq 2.$$

Now if we assume that a simple parallel capacitor configuration is used then the transformation that takes place is (see Fig. 11):

$$R_p = R_s(1+Q^2)$$
$$= 4{\cdot}9 \times 5 \simeq 25\,\Omega. \qquad (21)$$

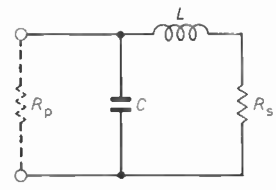


Fig. 11. Parallel capacitor configuration of device of Fig. 10.

Thus, by using a capacitor ($X_c \simeq X_L$) the impedance seen at R_p is real and approximately 25 Ω. If the input is required to match to 50 Ω a further LC section can be used with a very low Q, but the bandwidth is now almost entirely determined by the first section.

Assuming that a transformation of 50 to 5 Ω is required, then according to Matthaei[16] $r = 10$; if the required bandwidth

$$W = \frac{f_h - f_l}{f_m} = 1$$

then, at best one can achieve a 3·6 dB ripple for a simple transformation. In terms of the limited power gain available near f_t this is not often acceptable. A two-section filter would reduce this ripple to 1·6 dB whereas a three-section filter could minimize the ripple to 0·5 dB. It is therefore imperative, for large bandwidth and low ripple applications, that the Q is kept very low in the transformation section closest to the actual device bar.

Bringing the shunt capacitor closer to the device bar, in order to lower the Q, means that the interconnecting

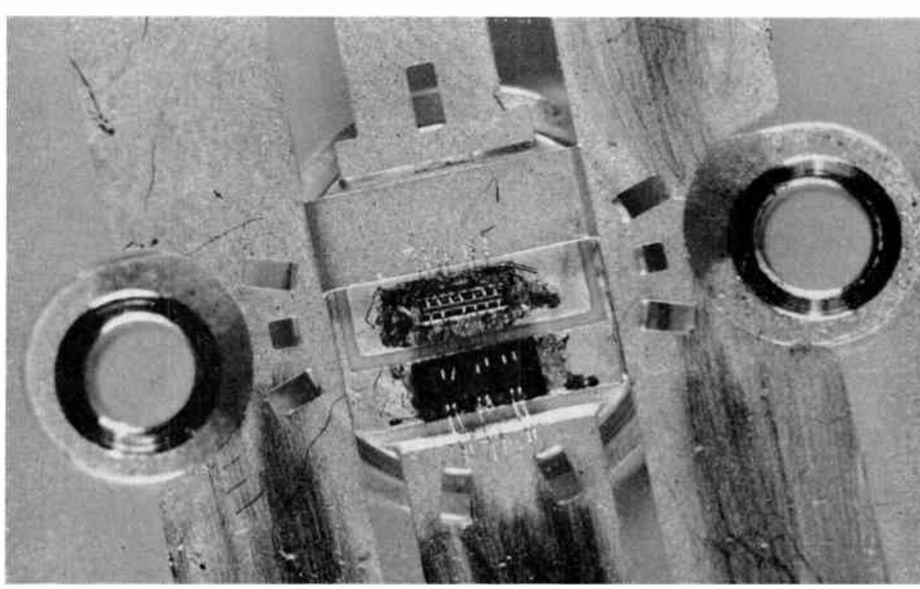


Fig. 10. Improved JO construction showing TI experimental device. Capacitor and bridging bar.

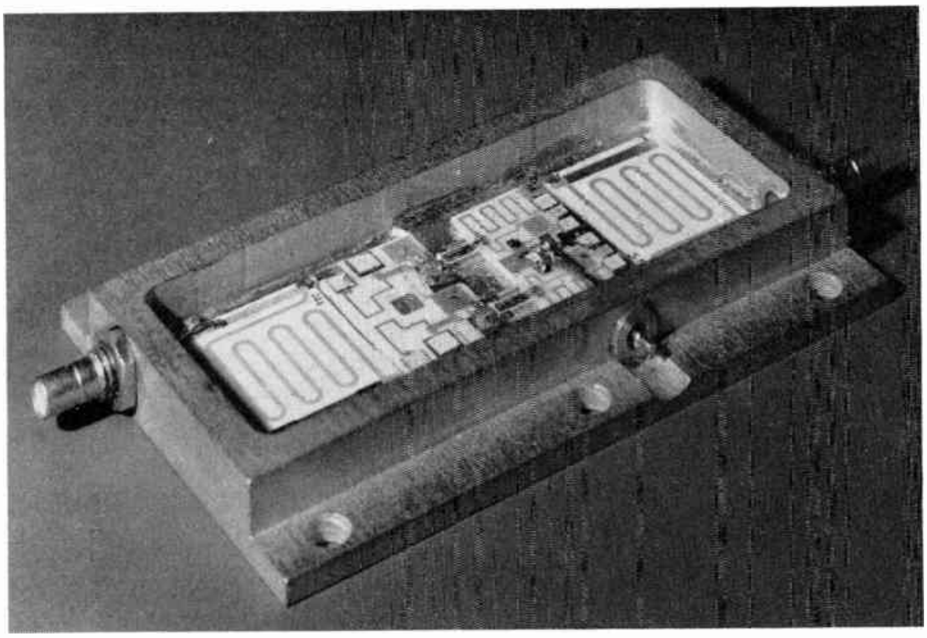


Fig. 12.
Wideband power amplifier.

value of L_b will fall and this implies that the value of the shunt capacitor must rise.

Rather than using multi-layer ceramic chip capacitors, most semiconductor manufacturers have found that the shunt capacitors can readily be made in silicon.

8 High Power Modules

Improvements in discrete power devices will continue, but the problems discussed in terms of bandwidth and power gain loss due to interconnexion is not unique to active devices, they also prevail in the circuitry. This can be well illustrated by a now well-established wideband power amplifier shown in Fig. 12.

The module consists essentially of two well-matched amplifiers, which are interconnected at the input and output by 3 dB couplers. As shown in equation (19) device gain is a function of frequency, as the modulus of the current gain varies with frequency. Thus if a 'flat' gain is required over the bandwidth, then this is generally achieved by increasing the mismatch towards the lower end of the band. In order to avoid the problems associated with a high v.s.w.r. two balanced amplifiers are used to obtain effectively conjugate matching to each other over the band.[17] Only the difference between the two is now absorbed in the terminating resistor to the 3 dB couplers. The input matching circuits consist of a three-section low-pass matching filter and the output circuit is a two-section filter. The two transistors are mounted on beryllium oxide. This construction largely eliminates some of the inductance problems associated with discrete devices. In the section of the input matching filter closest to the devices, the inductance problem arises again, in this case in terms of the inductance in series with the first chip capacitor. This can only be minimized by using thin film techniques. Instead of inductors small sections of line are used which yield an inductive impedance of the value of the characteristic impedance when the line is an eighth of the wavelength.

$$|X_L|_{\lambda/8} = z_0. \qquad (22)$$

This shows that terminating a large capacitor pad to earth will add the inductance of the pad in series with the chip capacitor. Using thin film technique the earth 'through connexion' can be made immediately under the capacitor chip. This amplifier (TPM 200/30) is designed for 225 to 400 MHz and at higher frequencies of course the problem becomes even more magnified. For power amplifiers in the telecommunication range of 1·8 to 2·2 GHz the distributed configuration shown in ref. 9 appears to be the correct answer, i.e. a modified, distributed Chebyshev transformer with capacitive stubs. The most outstanding electronic characteristic of thin film modules is the consistency that can be achieved and this is the main reason for the extensive use of thin film in the communication power modules where bandwidth is not a problem (the TPM 100 series). This is because consistency can be directly equated with the economics of sub-system assembly.

9 Conclusions

Enormous progress has been made since the transistor was invented but it is probably only during the last decade that the real impact of semiconductor technology has become apparent. With great rapidity, frequency and power achievements were superseded. In this review it has been shown that sophisticated technologies are now being introduced to conquer the next barriers. Improved mask making will allow finer structures to be used, shallow As diffusions will overcome emitter run-on, gold contact systems will improve the current handling capability as well as avoid emitter wash-out, emitter ballasting and collector cells improve the ruggedness of devices. At present, it looks as if the encapsulation of the active wafers will need to be improved fully to utilize the intrinsic device capability, but even this problem is being solved. The effect of lead inductance on power gain has been illustrated and some of the misconception on interdigitated structures discussed. In a rationale of the future it appears that for power applications thin film techniques and module construction offer the greatest chance of achieving further improvements.

Several authors have predicted what the future could have in store for us and of particular interest are some of the projections by J. D. Adams.[18]

If one however considers the Johnson limits mentioned earlier on, it appears that the useful limit of power in silicon is now in sight. Assuming that the Johnson limit of 2×10^{11} V s^{-1} is going to be realized one can then derive the maximum useful frequency for Si power. Again assuming that $f_{max} \simeq f_t$ and that the lowest useful battery voltage is say 12 V, then V_{max} is approximately 30 V and f_t or f_{max} is approximately 7 GHz. This means that other devices must take over from diffusion devices above these frequencies, at least for power generation. However, one must not lose sight of other factors either.

Communication efficiency is not solely a function of transmitted power, it is also a function of the sensitivity of receivers and in this field considerable progress is being made with GaAs Schottky barrier f.e.t.s. Noise figures, which in the end limit the receiver sensitivity, are beginning to fall rapidly now.

As GaAs f.e.t.s are just beginning to be developed into practical units, this area of progress may well dominate the scene during the next plan of communication progress.

10 Acknowledgments

The author would like to thank his colleagues for some useful discussions and help received, in particular Messrs. Garry Garrard, David Bonham and Derrick Robinson.

11 References

1. Brugler, J. S., 'Silicon transistor biasing for linear collector current temperature dependence', *I.E.E.E. J. Solid State Circuits*, **SC-2**, pp. 57–8, June 1967. (Letters.)
2. Runyan, W. R., 'Silicon Semiconductor Technology', Chapter 7. T.I. Electronics Series (McGraw-Hill, New York, 1965).
3. den Brinker, C. S., 'Die Untersuchung von Fehlerquellen integrienten Gleichspannungsvesstaerkern', *Arch. Elektrotechnik*, **51**, p. 5, 1968.
4. Early, J. W., 'Structure determined gain-band product of junction triode transistors', *Proc. Inst. Radio Engrs*, **46**, p. 1925, December 1958.
5. Beaufoy, R. and Sparkes, J. J., 'The junction transistor as a charge controlled device', *A.T.E. J.*, **13**, No. 4, pp. 310–27, October 1957.
6. Johnson, E. O. and Rose, A., 'Simple general analysis of amplifier devices', *Proc. Inst. Radio Engrs*, **47**, pp. 407–18, March 1959.
7. den Brinker, C. S., Fairbairn, D. and Norris, B. L., 'An analysis of the switching behaviour of graded base transistors', *Electronic Engng*, **35**, pp. 500–5, August 1963.
8. Gri, N. J., 'Microwave transistors—from small signal to high power', *Microwave J.*, pp. 45 *et seq.*, February 1971.
9. den Brinker, C. S. and Parkyn, M., 'Analysis of some wideband amplifier systems', *Design Electronics*, **5**, No. 8, pp. 34–9, May 1968.
10. Johnson, G., 'Power gain and stability in linear active two-ports', Chap. 4 in 'Solid-State Communications: Design of Communications Equipment using Semiconductors', T.I. Electronics Series (McGraw-Hill, New York, 1966).
11. Jacobson, D. S., 'What are the trade-offs in r.f. transistor design?' *Microwaves*, July 1972.
12. Archer, J., 'State of the art small signal microwave transistors', *Microwave J.*, February 1972.
13. 'Power transistors climb to 4 GHz with 5 W out.' (Cover feature.) *Microwaves*, March 1972.
14. Black, J. R., 'Etch pit formation in silicon at Al-Si contacts due to transport of silicon in aluminium by momentum exchange with conducting electrons'. Abstract, Electrochemical Society meeting Autumn, October 6–11th 1968.
15. 'A special report on transistor reliability', *Microwave J.*, July 1972.
16. Matthaei, G. L., 'Tables of Chebyshev impedance transforming networks of low-pass filter forms', *Proc. Inst. Elect. Electronics Engrs*, **52**, pp. 939–63, August 1964.
17. Engelbrecht, R. S. and Kurokawa, K., 'A wide band low-noise L band balanced transistor amplifier', *Proc. Inst. Elect. Electronics Engrs*, **53**, pp. 237–47, March 1965.
18. Adams, J. D., 'Capability of a projected 1975 airborne solid-state phased array radar', *Microwave J.*, September 1971.

Manuscript received by the Institution on 18th October 1972. (Paper No. 1495/CC159.)

UDC 621.375.049.774.621.396.62.049.77

Integrated circuits for analogue systems

Professor W. GOSLING,
B.Sc., C.Eng., F.I.E.R.E.*

SUMMARY

Because of the need for a large market and the initial problems with technically complex circuits, integrated circuits developed later for analogue than for digital systems. There are now two reasonably distinct groups of analogue circuits, namely those for professional use and those for incorporation in consumer equipment.

In the former category characteristics and trends of professional circuits are considered with particular reference to operational amplifiers and other types of amplifier, voltage regulators, analogue multipliers and phase-locked loops. Consumer integrated circuits considered include audio amplifiers, radio receivers and allied devices, and circuits for colour television.

In all cases there is a tendency for monolithic designs to diverge from established circuit practice in discrete technology, and this trend may be expected to continue.

* *School of Engineering, University College of Swansea, Singleton Park, Swansea, SA2 8PP.*

1 Application Trends

The monolithic integrated circuit technology achieved its first successes in the digital field rather than analogue because of many factors, not the least being that the rather simpler circuit topology came much sooner within the realm of feasibility for production at acceptable yield levels, and that the circuits could operate well with rather large device parameter tolerances. Linear circuits are more exacting technically and consequently developed later.

Quite as important, however, was the size of the immediately accessible market. Even a single computer uses very large numbers of rather similar circuits and before the emergence of monolithic technology computer designers were in the habit of designing around standardized gates. Thus at the beginning of the 'sixties it was apparent that a monolithic gate could achieve acceptance in a market of millions or tens of millions of units per annum. As a consequence substantial resources were devoted to the development of a standard logic system, with t.t.l. emerging as the currently accepted industry standard. That this technical triumph was mishandled by the industry in a manner reminiscent of the economics of the 'pig cycle' is regrettable, but not germane to the present contribution.

It was natural that attention should be given to finding, in the field of linear circuits, an operational block, or sub-system, which could possibly also sell to users in very large volume. The need for the integrated circuit industry to operate on a large scale of production in order to achieve the full cost-saving potentialities of the technology has been explained so many times in the literature that it would be tedious to re-iterate it here. It costs only a little more to make an integrated circuit than to make a single transistor provided that the number made is large enough to spread the high design and mask production costs over a large number of units.

The first approach to the linear market was through monolithic operational amplifiers, and the first truly successful linear integrated circuits were the 702 and 709 operational amplifiers, and the 710 and 711 comparators, all designed by R. J. Widlar. These devices have been made in large volume and are still in widespread use, although the first two must now be regarded as obsolete. Most of the disadvantages of the 709 were eliminated in the 741, which replaced it—a device of great sophistication now in the very widest use and sold at a price little more than one-fiftieth of that at which a discrete amplifier of similar specification might have been sold ten years before. To a lesser degree the 741 has had a commercial history which parallels that of t.t.l., overproduction leading to very low prices; however, in this case the semiconductor manufacturer has suffered less severely because using 741's has encouraged designers to use other linear i.c.s, on which the manufacturer secures a more favourable return.

The operational amplifier was not the only general purpose i.c. to be put forward. An early contender for large-scale markets was a class of circuit containing an array of diodes or transistors or both with minimal interconnexion, intended for direct incorporation in discrete component linear circuits in place of a number of independent devices. The principal advantage put forward for array circuits of this type was space and possibly cost saving relative to discrete transistors; it was also suggested that occasionally the match between devices on the same chip could be exploited. Because in designing the more complex analogue systems it is virtually always necessary to combine a certain amount of discrete construction with the monolithic circuits, transistor arrays will continue to find applications. They have not, and will not, take a major share of the total of linear i.c.s used because they are too simple to be optimal.

The cost of an integrated circuit is only weakly dependent on its complexity; thus the advantage over a discrete approach increases sharply as the complexity of the function performed increases. An array circuit which gives four, or perhaps six, transistors for the cost of, say, two discrete devices does not represent nearly so good a bargain as, say, a 741 which contains two dozen transistors and many more passive components at much the same price. Thus the outstanding problem of linear integrated circuits, from the point of view of the manufacturer anxious to establish an economically viable production line, is the production of circuits of a very high order of complexity which will nevertheless sell in large volume. The normal solution is to identify relatively complex functions for the circuits to perform, and to place specifications on them which would be difficult to meet using a discrete component approach at anything like comparable cost.

Thus an operational amplifier not only provides a large gain, it also acts as an impedance transformer and a double-to-single-ended converter. It is a device of formidable complexity in discrete component form, yet easily fabricated in monolithic form.

Two distinct patterns in the development of linear i.c. technology are now discernible (with some overlaps), corresponding to the professional equipment and consumer durable markets. The former naturally developed first, but, particularly as a result of the cut-back in defence electronics expenditure in the USA, suffered something of a setback to its growth in 1970–71. At that time the sale of colour television receivers was very substantial in the US and growing in Europe, and i.c. designers turned their attention ever more strongly to this potentially very large market, well suited to i.c. technology by virtue of the standardization of the circuit functions to be performed and the very large volume of television receiver production.

Operational amplifiers had been the starting point for the professional linear integrated circuit, later joined by other circuits, such as power supply regulators, analogue multipliers and video amplifiers, able to perform identifiable circuit functions of a relatively simple kind. By contrast the complexity of circuits for the consumer durable market has been substantially greater. The first circuits of this type to appear were radio receivers, audio amplifiers, and television sound systems. Although, for example, phase-lock loop circuits may be used in both spheres, generally the two areas of circuit development have remained distinct, and will therefore conveniently be described separately in what follows.

2 Linear I.C.s for Professional Equipment

As already indicated, the first widely-used professional linear i.c. was the operational amplifier, introduced commercially in 1963. Devices of this type still have the highest sales of any linear circuit, but they are followed by voltage regulators, now available at low cost and in extensive use. Circuit blocks for use in radio receivers and similar applications were also available at an early date. These include mixers, gain blocks intended as a.g.c. r.f. and i.f. amplifiers, demodulator and a.g.c. generators, and so on. More recently analogue multipliers have been increasingly widely used, and phase-lock loops, for use in tracking filters and synchronous demodulators, have been introduced. More specialized circuits, such as voltage to frequency converters, are also available.

2.1 *Operational Amplifiers*

The variety of operational amplifiers now available is large, perhaps bewilderingly so. The general purpose device now mostly widely used is the 741. First described by Fullagar[1] and Widlar,[2] it quickly replaced earlier comparable devices, such as the 709, perhaps most of all because it was much easier to obtain stable circuit operation. Even with external frequency compensation, earlier devices frequently oscillated when the circuit layout adopted was unfavourable. They also had the reputation of electrical fragility, overcome in the 741 by the incorporation of circuits which limited the output current even under short circuits from the output terminal to either supply line. Early operational amplifiers also suffered from an irritating inversion of the sign of the gain for large input overloads, with the consequence that feedback in biasing circuits turned from negative to positive, causing the amplifier to 'latch up' stably with the output equal to its maximum excursion. To overcome these problems a 709 amplifier would be associated with two 'anti-latch-up' diodes at the input, a frequency compensating resistor-capacitor network, a current limiting resistor in series with the output, and one or two by-pass capacitors on the supply lines to suppress high frequency oscillation. By contrast the 741 may be used in similar applications with no additional components at all.

Although the 741 is available at very low cost, in a variety of packages and in dual form, a range of other amplifiers is also produced which meet needs beyond the capabilities of this excellent device, although often at a penalty in respect of some other aspect of its specification. There has, for example, been considerable interest in amplifiers having a substantially faster slew-rate than the modest 0·5 V/μs of which the 741 is capable. A well-established design permitting a slew-rate of 18 V/μs is the μA715 described by Narayanamurthi.[3] Higher slew rates are achieved principally by the use of emitter degeneration in the early stages of the amplifier avoidance of lateral p-n-p devices, and by the use of emitter followers combining an n-p-n device with the unusual feature of a vertical p-n-p, at the output. However, external frequency compensation will be needed with this design if feedback is used.

Similarly, the input impedance of the 741 is 1 MΩ, but for some purposes a much higher figure is essential. Considerable attention has been given to operational amplifiers with f.e.t. input stages, and for a time hybrid types, with the f.e.t. pair not included on the main amplifier chip but connected via a thick film circuit, were commonplace. Now monolithic f.e.t. input amplifiers are proving satisfactory, and, for example, the type 8007 has an input resistance of 1 terohm (1 TΩ = 10^6 MΩ), 1 pA input current, and a slew rate of 6 V/μs whilst retaining internal frequency compensation.

High voltage (± 40 V) operational amplifiers, devices capable of delivering up to 1 A to a load, and micropower amplifiers capable of operation with a total standby power measured in nanowatts are all available to the linear system designer.

Some micropower amplifiers are examples of an interesting circuit approach which differs quite radically from that normally adopted and pioneered by Widlar with the 702.[4] The conventional circuit uses a differential input stage, of low-drift, high-input impedance and good common-mode-rejection, followed by a stage providing double-to-single-ended conversion (with level shift) and finally an output stage giving a low output impedance. Most operational amplifiers have more or less followed this pattern, but a different approach is presented by the family of operational transconductance amplifiers (o.t.a.) first described by Wheatley and Wittlinger.[5] This type of amplifier is distinguished by having an output stage of very high output impedance. As a result a circuit is possible (Fig. 1) which uses only active devices, and no

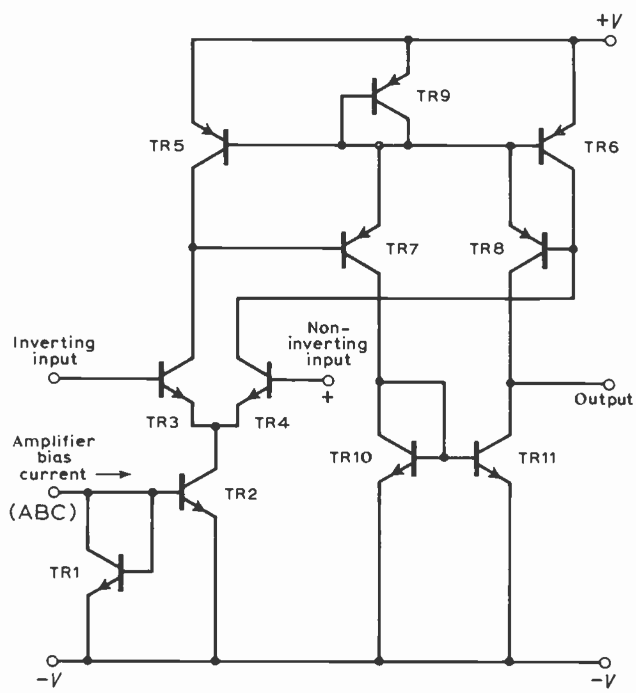


Fig. 1. Basic o.t.a. circuit (*Courtesy of RCA Solid State*).

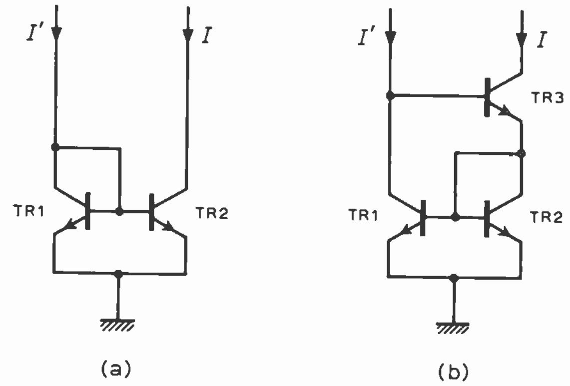


Fig. 2. Current mirrors.

resistors. Before considering the operation of this circuit, consider the transistor combinations of Fig. 2. These are known as 'current mirrors' and form the basis of the o.t.a. design. In Fig. 2(a) the current I' flowing through TR1 (which is diode connected) and the base of TR2 establishes a voltage across TR1 which is also the base-emitter voltage of TR2. Assuming that the two transistors are identical, they will have the same collector current, hence

$$I = I' - 2I_B$$

where I_B is the base current or

$$I = \frac{I'}{1+2/h_{FE}}$$

to good approximation.

Since h_{FE} is likely to be 100 or more, I is very little less than I'. This simple circuit only gives approximately equal I and I' because increase of collector voltage applied to TR2 will cause an increase in I as a result of its relatively modest output resistance. The circuit of Fig. 2(b) improves the constancy of the current I by applying current derived negative feedback to TR3 via TR2 and TR1. Also the emitter current of TR3 is I_E, where

$$I_E = I + I_B$$

If the collector current of TR1 and TR2 is I_C,

$$I_C = I' - I_B$$

The emitter current of TR3 is equal to the collector current of TR2 plus $2I_B$, hence

$$I_E = (I' - I_B) + 2I_B = I' + I_B$$

Thus

$$I = I'$$

In this case the approximation is very close, and, to a first-order, not dependent on h_{FE}.

Referring to Fig. 1, it will be seen that TR1 and TR2 form a current mirror which provides the 'tail' current for TR3 and TR4, which form a 'long-tailed pair' (emitter-coupled) amplifier. Because the sum of the collector currents of TR3 and TR4 is constant (equal to the tail current), the voltage across TR9 is constant, resulting in constant collector current through TR5 and TR6, which is arranged to be slightly less than the collector currents of TR3 and TR4. The difference between the constant current through TR5 and the collector current of TR3 is delivered as base current to TR7, where it is amplified by h_{FE}, and similarly TR8 amplifies the output of the other side of the 'long-tailed pair'. A current mirror TR10, TR11 ensures that the current delivered to the output terminal is equal to the difference between the collector currents of TR7 and TR8, and hence h_{FE} times the difference between the collector currents of TR3 and TR4.

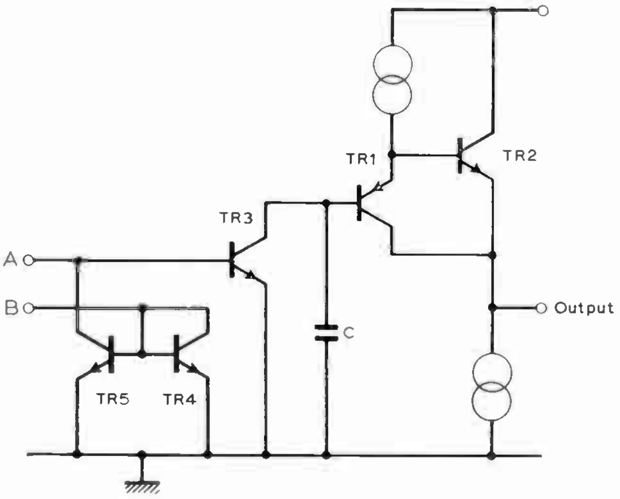


Fig. 3. A current differencing amplifier.

Practical o.t.a. circuits contain additional transistors principally to overcome the effects of poor device parameters in lateral p-n-p devices and further to raise the output impedance. All however retain the feature that the circuit is composed of transistors only, without resistors, and that the current through every transistor is proportional to the emitter current of the input stage. The device is thus equivalent to a transconductance of a value set by the value of the emitter current. Applications are to variable gain amplifiers, gated amplifiers and modulators as well as to the more normal fixed gain applications. By choosing a very low value of emitter current micropower operation is possible. A simple post-amplifier (using a f.e.t.) will provide a low output impedance amplifier if so required.

The possibilities of o.t.a. devices have not yet been fully exploited by system designers, but obviously in certain respects they confer freedom of action which could be exploited to advantage. Another new circuit is the current differencing operational amplifier described by Frederiksen, Davis and Zobel,[6] who emphasize its ability to operate from a single polarity supply unlike the conventional operational amplifier, which requires supplies of both polarities. Referring to Fig. 3, the inverting input terminal is A and the amplifier will be seen to consist of a common emitter stage followed by two cascaded common-collector stages, to give a high voltage gain and low output impedance. The provision of a non-inverting input terminal, to give the usual differential input exploits the properties of the current mirror formed by transistors TR4 and TR5. A current flowing into terminal B results in an almost equal current flowing out of terminal A to the collector of TR5. Thus the current flowing to the base of TR3 is equal to the difference of the currents flowing into terminals A and B. The capacitor C (some 3 pF) gives the amplifier a first-order high frequency response.

The main attraction of the amplifier is its simplicity, and hence small chip area, resulting in low manufacturing cost. Four dual input amplifiers can conveniently be incorporated in a single 14-lead d.i.p. The unconventional input circuits, used with input resistors, give a common-mode range effectively limited only by the current ratings of TR4 and TR5, and hence very large.

The conventional operational amplifier is already a very sophisticated circuit element: new developments along the lines described promise continuing extension of the range of its usefulness.

2.2 *Other Amplifiers*

A variety of integrated circuit amplifiers for use in narrow-band i.f. and r.f. amplifiers are available. They mostly feature means of gain control for a.g.c., and usually have well-defined input and output conductances which facilitate the design of stable amplifiers. They are not, however, at present competitive with discrete component amplifiers in respect of ultimate noise or linearity performance. An exception is the dual-gate m.o.s. transistor, which may be regarded as an integrated m.o.s. cascode amplifier, and is the best low level r.f. amplifier currently available up to about 400 MHz. By contrast, there are few bipolar integrated r.f. amplifiers which can be operated beyond 100 MHz.

A variant of some interest is the wideband limiting amplifier, usually consisting of a series of cascaded emitter-coupled pair stages ('long-tailed' pair), which limit symmetrically by virtue of the symmetry of their topology. An example is type TAA350 (Fig. 4) which limits any input in excess of 100 μV amplitude, and is usable to beyond 15 MHz. Limiting amplifiers are often combined with f.m. demodulators, and the properties of such a combination will be described later.

Audio power amplifiers are well established as monolithic devices, but their properties will be discussed below, in connexion with consumer applications.

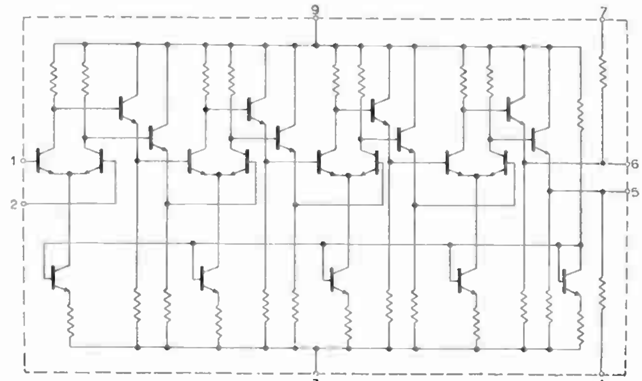


Fig. 4. A limiting i.f. amplifier, type TAA 350 (*Courtesy of Mullard*).

2.3 *Voltage Regulators*

Because power supplies are standardized at 5 V for logic circuits and ±15 V for linear i.c.s, voltage regulators are potentially capable of providing the large market for a standardized product needed to support the successful introduction of integrated circuits. In recent years sales of integrated voltage regulators have justified this expectation, rising to take second place only to operational amplifiers and comparators. They have done so, however, partly as a result of a change in the philosophy of power supply design.

Until quite recently it was commonplace to use only a single regulated supply at each required voltage level even in complex electronic systems, distributing the supply to individual printed circuit cards via impedance bus-bars, which were intended to have negligible voltage drop and thus also not to introduce unwanted coupling between circuits. With the availability of cheap i.c. regulators, this philosophy is being abandoned in favour of the distribution to the printed circuits of unregulated or, at most, roughly regulated supplies and using an i.c. regulator of more modest current rating on each individual card. As a result emphasis in regulator design has passed to relatively simple-to-use devices needing an absolute minimum of external components and able to supply the standard voltages.

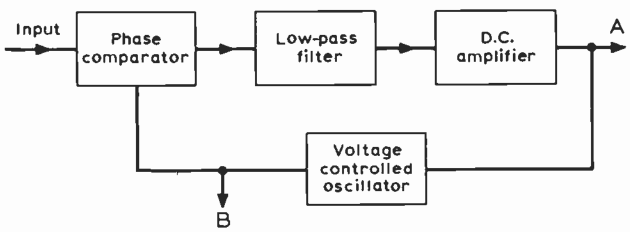


Fig. 5. A phase-locked loop.

A good example is the 5 V LM109 regulator described by Widlar[7] which has only only three terminals (unregulated input, regulated output, common). Capable of an output of 1·5 A and with an output impedance of only 30 mΩ, this device gives a line regulation of 50×10^{-6}/V. Earlier i.c. regulators relied on sensing excessive current to protect the regulator from the effect of short circuits, but this indirect mode of protection depended on the use of a suitably chosen external current sensing resistor, since the actual maximum permissible current depended on extent of heat sinking, supply voltage and other factors. The regulator described by Widlar is protected in a more satisfactory and fundamental way by detection of temperature rise in the series transistor, using the effect of temperature on the characteristics of a transistor mounted adjacent to the series transistor. Current limiting and crowbar clamping then play a purely secondary protective role.

A similar temperature rise sensing approach to protection is described by Davis[8] for a ±15 V five-terminal regulator for linear i.c.s. Regulators of this kind must demonstrate excellent tracking between positive and negative supplies, which demands careful attention to the symmetry of the layout, so far as temperature distribution is concerned. Also, rather low noise levels are required on the regulated supplies to high-gain operational amplifiers, thus careful attention is given to the noise output associated with the Zener voltage reference.

On-card regulation is now becoming virtually a standard technique, and the development of simple fixed voltage regulators is exerting pressure on designers of other i.c.s to adhere to the standard 5 V and ±15 V supplies. Thus the regulator market is likely to continue to exhibit growth.

2.4 *Analogue Multipliers*

It was Gilbert[9] who demonstrated that a precision analogue multiplier, based on a development from the emitter-coupled pair circuit, could be constructed in monolithic technology. A device of this type is the MC 1596 which, used as a balanced modulator, achieves carrier suppression of as much as 66 dB. Circuits of this type may be used as synchronous demodulators, suppressed carrier modulators, frequency and phase discriminators and in analogue computing circuits for multiplication, division, squaring and so on. A disadvantage is that the output is floating, and thus a double-to-single-ended converter must follow. Want of balance in this circuit can result in a serious deterioration in the apparent performance, for example in a balanced modulator carrier suppression deteriorates sharply. More recent circuits, such as the CA 3091D, give a single-ended output, which is usually more convenient. Future developments in these circuits will probably include universal adoption of single-ended inputs and outputs and extension of the frequency response into the higher v.h.f. range.

2.5 *Phase-locked Loops*

Because the cost of a component is proportional to the area that it occupies on a silicon chip, monolithic circuits cannot incorporate inductors. Thus much attention has been given to the design of integrated subsystems which are capable of performing functions which in discrete circuits would depend on the incorporation of inductors. An important example is the phase-locked loop, shown in block diagram form in Fig. 5. When the frequency of the incoming signal is sufficiently close to the free-running frequency of the voltage-controlled oscillator an output is obtained from the phase comparator which will pull the v.c.o. into synchronism with the incoming signal. When locked in, the v.c.o. gives an output at point B which is of the same frequency as the input and in phase quadrature to it. The voltage at point A is proportional to the difference between the v.c.o. free-running frequency and the input frequency.

Thus taking the output from B, the phase-locked loop (p.l.l.) may be regarded as equivalent to a tracking limiting filter, whilst the output at A is the equivalent to that from a frequency discriminator, and can thus be used to demodulate f.m. signals.

It is the latter application to which most attention has been given hitherto, since the integrated p.l.l. can function as an i.f. amplifier, limiter and discriminator without the use of a single inductor. Grebene[10] has described the use of a monolithic p.l.l. for both f.m. and a.m. demodulation.

The phase comparator circuit is similar to the circuits described by Gilbert.[9] An ingenious multivibrator using a single external timing capacitor is used as the v.c.o. A differential circuit topology is used and the free-running frequency is made only slightly temperature dependent ($0{\cdot}06\%\,\text{deg}^{-1}$ C). In the circuit described (type NE561B) an additional multiplier is included to give the circuit a.m. as well as f.m. demodulating capabilities. An output taken from point B is multiplied with the a.m. input shifted by 90° in phase (for which purpose a simple CR network is all that is required). The p.l.l. locks to the carrier, and hence synchronously demodulates the a.m. input. Thus, apart from the v.c.o. tuning and low-pass filter capacitors, the complete a.m./f.m. demodulator is fabricated on a single chip. In practice the circuit can be arranged to be highly selective, for example when operating at 10·7 MHz with an f.m. signal of 75 kHz deviation, interference rejection of some 30 dB can be obtained for signals 200 kHz off from the wanted signal, which is said to be comparable with the selectivity normally obtained with a three-stage i.f. strip of conventional design. Full f.m. limiting occurs for inputs above 200 μV, for which an audio output of about 100 mV is obtained.

In addition to acting as an f.m. or a.m. demodulator, the p.l.l. has important applications in tracking filters, using filter circuits of the Barber type,[11] in f.s.k. modulation and demodulation, and in frequency synthesis.

The p.l.l. is an example of an integrated device which performs circuit functions in a manner essentially different from that commonplace in discrete circuit practice. Probably many other such circuits will evolve. Already other specialized monolithic i.c.s for professional applications are available and only considerations of space preclude their consideration here.

3 Integrated Circuits for Consumer Applications

Very many integrated circuits designed to be embodied in equipment intended for direct sale to the public have been launched on the market. Only a very brief review of them can be given here, touching on the points of particular circuit interest.

3.1 *Audio Amplifiers*

Amongst the first consumer-oriented integrated circuits to be released were audio amplifiers. These are now available at highly economic cost for power output ratings up to a few tens of watts. The most interesting circuit development has been the use of temperature rise protective circuits similar to those already described in connexion with supply regulators. By this means the device is protected against even the most severe misuse, although some additional overvoltage protection is also desirable. There is really little of technical interest to say about audio amplifier i.c.s which have taken on the character of a stable and developed technology, with little prospect of technical revolutions but with continuing and steady evolution towards improved device specifications.

3.2 *Radio Receivers and Allied Devices*

Radio receivers hold out promise of very large markets and are consequently attractive to the i.c. manufacturer. Both a.m. and f.m. receiver circuits have been marketed, but at present require large numbers of additional external components, including r.f. and i.f. filters, and consequently do not show the full economic advantages of i.c. technology. Possibly only when a radically new receiver configuration is used in place of the conventional superheterodyne circuits will this difficulty be overcome. The p.l.l. circuit, described in Section 2.4, makes possible an f.m. or synchronous a.m. receiver using audio low-pass instead of i.f. band-pass filters, but cannot yet in all respects compete with the conventional superheterodyne so far as its specification is concerned. Interest in direct conversion receivers, particularly for s.s.b., has been re-awakened, and, for example, the receiver described by Al-Araji and Gosling[12] is intended for integration in monolithic form with only a simple r.f. filter and local oscillator circuits external to the chip.

Although complete receivers are only slowly finding acceptance the same would not be true of two other fields: stereo decoders and television sound sub-systems. An example of the former is the type CA 3090Q, and its operation will be reviewed briefly.

The stereo decoder[13] must reconstruct the 38 kHz carrier suppressed in forming the stereo signal. It does this by selecting amplifying and doubling the 19 kHz pilot tone, all without disturbing its correct phasing relative to the 38 kHz $L-R$ difference signal information. Normal LC filters need to be very accurately tuned to perform the selectivity function without unacceptable phase shift. Here again, a phase-locked loop is the solution adopted. A 76 kHz voltage-controlled LC oscillator is counted down to 38 kHz by a symmetrical counter circuit, ensuring unity mark/space ratio in the carrier so generated (essential if optimum channel separation is to be achieved). Further division to 19 kHz permits phase comparison with the received pilot tone and hence phase locking of the v.c.o. The integrated circuit is also capable of performing a number of other important functions (Fig. 6 shows a block diagram). These include a pilot presence detection circuit, which lights a signal light and also brings the $(L-R)$ detector circuit into operation, the $(L-R)$ detector itself and the matrix by which the L and R channels are reconstructed from the $(L-R)$ and $(L+R)$ signals. A full circuit is shown in Fig. 7, which serves to illustrate the complexity of modern integrated circuits for consumer application. This approach to the design of a stereo decoder seems to be gaining fairly widespread acceptance.

So far as television sound integrated circuits are concerned, available circuits are compared rather fully by Blaser and Long.[14] The comparisons are very exhaustive and could not even be summarized here. Looking at the circuits in order of date of development the noticeable feature is the progressive evolution away from the circuit approaches used in discrete component

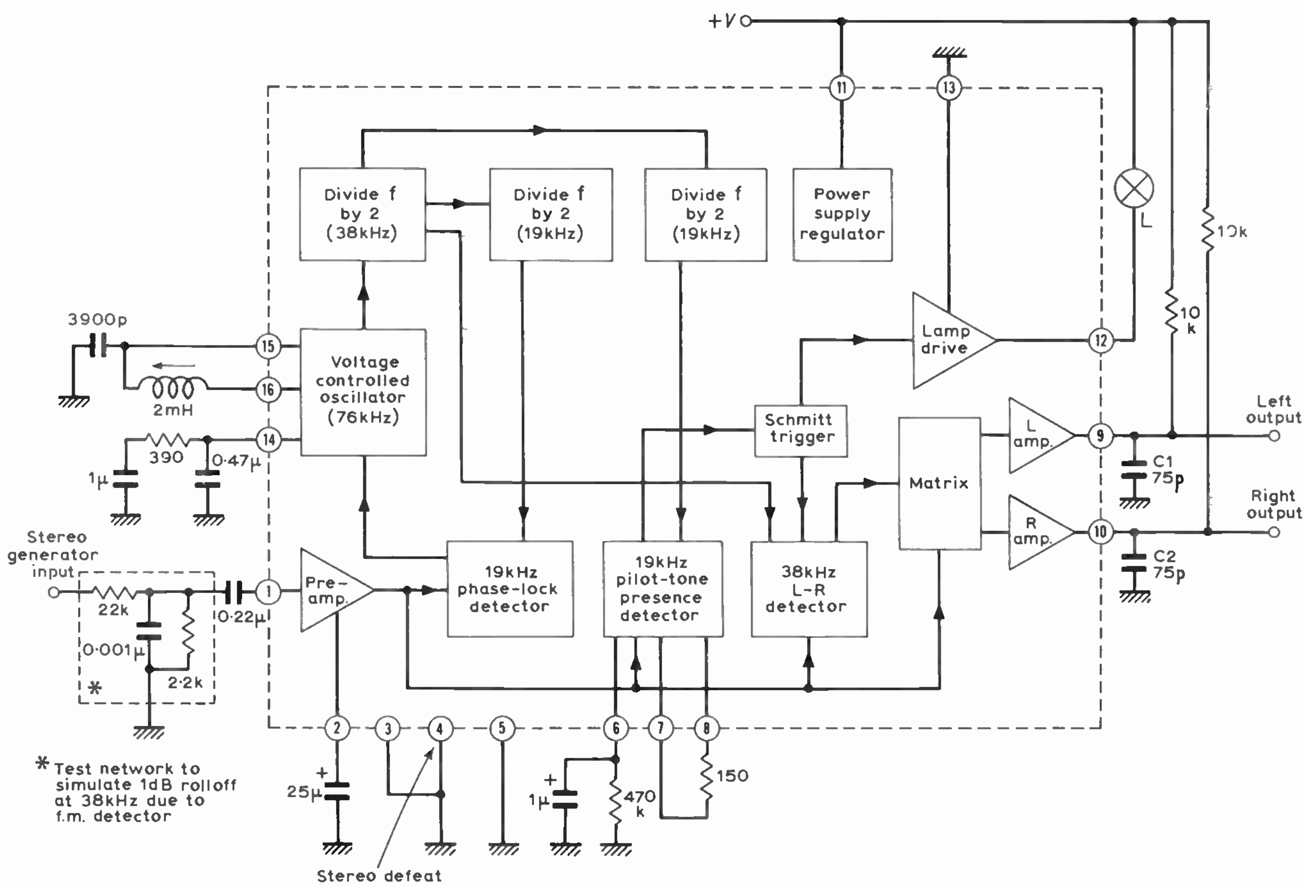


Fig. 6. Functional block diagram of the CA3090Q, phase-locked stereo decoder (*Courtesy of RCA Solid State*).

technology towards circuits particularly suited to monolithic fabrication. For example the Foster-Seeley f.m. discriminator used in early circuits gives way to a demodulator which relies on the phase shift in a single resonant circuit, with phase comparison made by a linear modulator of balanced type, as described above. For this application too, p.l.l. circuits show promise of a solution without any use of inductors at all.

3.3 *Circuits for Colour Television*

The most complex analogue integrated circuits at present available are some of those intended for use in colour television. Because the circuit functions to be performed are highly standardized the market is vast, particularly when, as is usually the case, circuits can be adapted to either PAL or NTSC and any of the line standards. Thus monolithic circuit manufacturers are encouraged to develop and market highly specialized and complex integrated circuits. For example the TBA 540 reference combination designed for the PAL system contains 57 transistors, and the TCA 270 synchronous demodulator some 93. Integrated circuits are available for virtually every function in the receiver except the u.h.f. front end, and indeed there is now a manifest tendency towards the 'all-i.c.' television receiver. Circuits which cannot be included in i.c.s designed to perform major sub-system functions are brought together in so-called 'jungle circuits' which, in effect, deal with all circuit functions not otherwise covered.

4 Conclusions

Integrated circuits for linear applications are daily becoming more numerous, cheaper, and of more advanced specification. The problems of achieving adequate volume of sales are being met by exploiting ever more fully the unique advantage of integrated over discrete circuits, namely the very low marginal cost of increased complexity. As a result system designers are learning that by using an integrated circuit where once they had used a few discrete transistors they may be able to improve performance and give themselves a dimension of design flexibility not otherwise possible. As a consequence a tendency is now quite clearly developing to fit system designs to available integrated circuits, thus further increasing the size of the i.c. market.

The observant will appreciate that this amounts to positive feedback: it has given rise to a rapid swing from discrete to monolithic technology in analogue systems which will accelerate until some natural limit is reached. In the author's opinion this limit is likely to be the effective elimination of discrete analogue circuits.

5 Acknowledgments

My thanks are due to my many friends in the semiconductor industry for advice on the content of this paper, and in particular to Mullard Limited and to RCA Solid State for permission to reproduce copyright material.

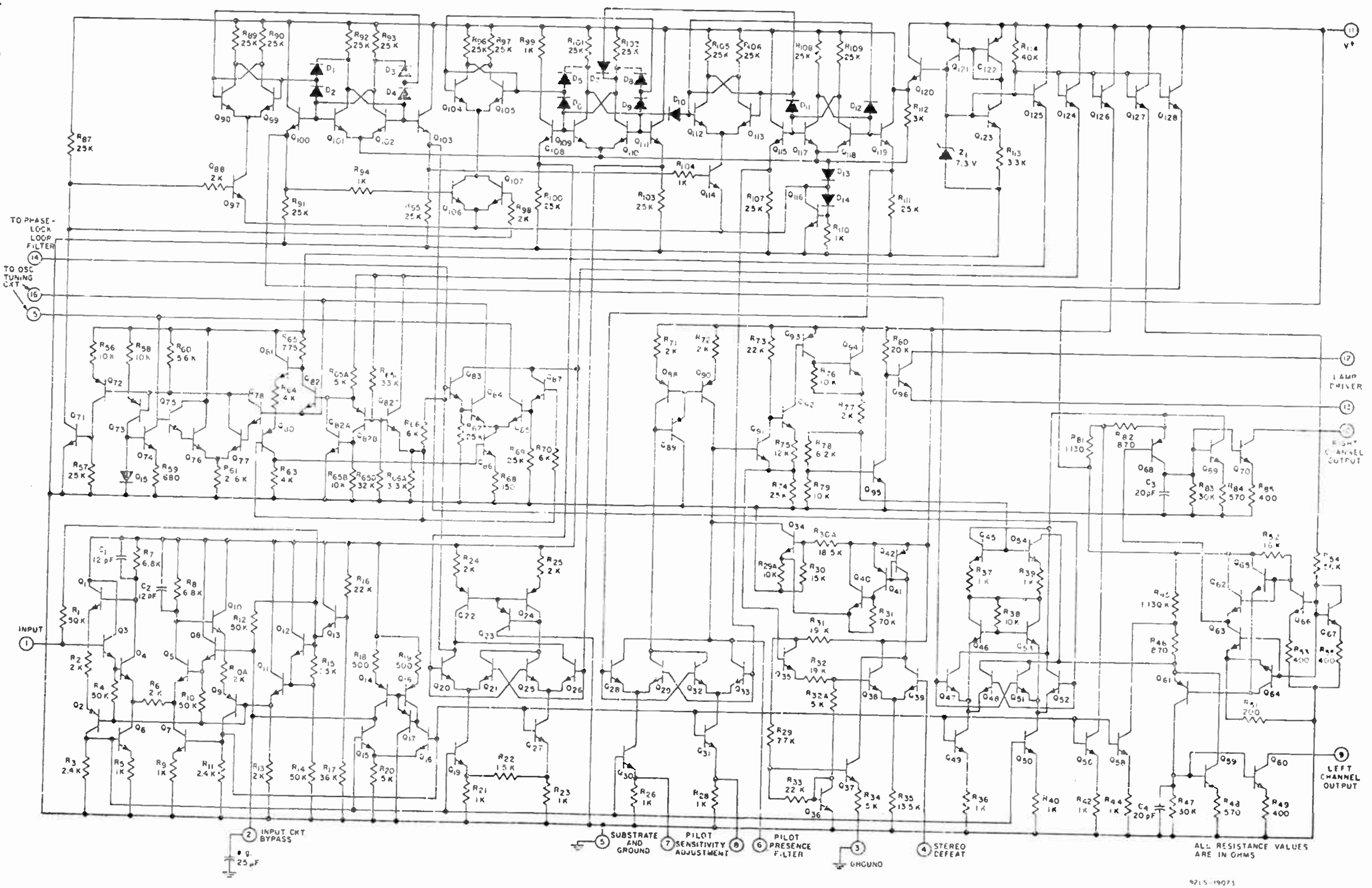


Fig. 7. Circuit schematic diagram of the CA3090Q. The size of the chip is 2·67 mm × 2·49 mm (105 × 98 mils) (*Courtesy of RCA Solid State*).

6 References

1. Fullagar, D., 'A new high performance monolithic operational amplifier', Technical Report, Fairchild Semiconductor.
2. Widlar, R. J., 'A new monolithic operational amplifier design', Technical Report TP2, National Semiconductor Corp., 1967.
3. Narayanamurthi, E. S., 'New high speed monolithic operational amplifier', *IEEE J. Solid State Ccts*, **SC-6**, pp. 71–76, 1971.
4. Widlar, R. J., 'Design techniques for monolithic operational amplifiers', *IEEE J. Solid State Ccts*, **SC-4**, pp. 184–91, 1969.
5. Wheatley, C. F. and Wittlinger, H. A., 'OTA obsolete op. amp', Technical Report ST 4159, RCA Electronic Components, 1969. (Reprinted from 1969 *NEC Proceedings*.)
6. Frederiksen, T. M., Davis, W. F. and Zobel, D. W., 'A new current differencing single supply operational amplifier', *IEEE J. Solid State Ccts*, **SC-6**, pp. 340–47, 1971.
7. Widlar, R. J., 'New developments in i.c. voltage regulators', *IEEE J. Solid State Ccts*, **SC-6**, pp. 2–7, 1971.
8. Davis, W. F., 'A five-terminal ± 15 V monolithic voltage regulator', *IEEE J. Solid State Ccts*, **SC-6**, pp. 366–76, 1971.
9. Gilbert, B., 'A new wide band amplifier technique', *IEEE J. Solid State Ccts*, **SC-3**, pp. 353–65, 1968.
10. Grebene, A. B., 'An integrated frequency selective am/fm demodulator', *IEEE Trans. Broadcast & Telev. Receivers*, **BTR-77**, pp. 71–79, 1971.
11. Barber, N. F., 'A narrow band filter using modulation', *Wireless Engineer*, **24**, pp. 132–4, 1947.
12. Al-Araji, S. R. and Gosling, W., 'Direct conversion s.s.b. receivers: a comparison of possible circuit configurations for speech communications', *The Radio and Electronic Engineer*, **43**, 1973, to be published.
13. Kaplan, L. A., Kleinman, H. M. and Limberg, A. L., 'Integrated circuit decoder does everything', *IEEE Trans. on Broadcast & Telev. Receivers*, **BTR-17**, pp. 202–12, 1971.
14. Blaser, L. and Long, D., 'A comparison of integrated circuit television sound systems', *IEEE Trans. on Broadcast & Telev. Receivers*, **BTR-17**, pp. 35–42, 1971.

Manuscript received by the Institution on 29th November 1972. (*Paper No. 1496CC160.*)

UDC 621.382 : 621.374

Trends in semiconductor digital circuits

K. J. DEAN,
M.Sc., Ph.D., F.Inst.P., C.Eng., F.I.E.E., F.I.E.R.E*

SUMMARY

The ways in which digital circuit design techniques have developed over the last 25 years are described. The interaction between circuit design and devices is discussed in the light of the range of compatibility of digital circuits and the systems which became possible with semiconductors which were not so with thermionic valve circuits.

* *Principal, South East London Technical College, Lewisham Way, London, SE4 1UT*

1 Immediate Post-war Developments

When engineering history comes to be written from a more distant viewpoint it may be that special significance will be seen in the developments, almost coincident in time, of semiconductors, electronic computers and nuclear engineering. That is not to say however that the first computers used semiconductors, other than a few d.c. restoration diodes. The *Pilot Ace* designed and built at the National Physical Laboratory in 1949–52 used ECC 82 double-triodes and Z77 high-slope pentodes in a cathode-coupled logic circuit. (Nowadays it would have been called 'c.c.l.' but then 'drop-of-a-hat' abbreviations were not the fashion.) The contribution of computer design to electronics, even at the start, was the emphasis it placed on the concept of modular design, based on the idea of a logical gate. In those days this was in the context of pulse and switching circuits and later of 'static switching'; but although these terms are no longer in favour this concept of functional blocks has always been particularly clear in that area. The early computer designers were digital circuit engineers who appreciated from the start the repetitive structure of interconnected gates, even though the poor reliability of vacuum devices set a limit on the size and complexity of practical systems using that technology. To-day with semiconductor technology that limit has been pushed far back.

War-time radar had introduced a generation of engineers to Eccles-Jordan flip-flops, *kipp* relays and triggers: all thermionic digital circuits. The nuclear explosions which brought World War II to such a sudden and terrible halt also laid the foundations of nuclear instrumentation, as nuclear physics was developed for scientific research and for power generation. Coincidence counters were really AND-gates and anti-coincidence counters used exclusive-OR circuits. The need for general counting circuits, pulse height discriminators and later digital frequency meters was accompanied by the development of thermionic and cold cathode special purpose tubes such as the dekatron. The contribution of nucleonics was therefore to extend the field of the digital circuit designer outside computing into new areas of instrumentation.

Into all of this came the transistor: small, with reasonable voltage levels and modest power requirements and potentially not so modest reliability. What problems it presented were ones of amplifier linearity and high power and high frequency operation. It was surely no wonder that it was in digital circuitry that transistors and later integrated circuits have made so much impact.

2 A Variety of Gates

There were many examples of logical systems before engineers began to think in terms of electronics. Caldwell[1] mentions the railway interlocking signal system of the nineteenth century, there were relays, multi-way lighting systems and many others. Thermionic equipment was often inappropriate, costly, bulky and unreliable. Transistors changed all that. Development was at first slow. Industry for the most part waited on the Bell symposium in 1952 at which most of the major companies were represented. However the licensing arrangements

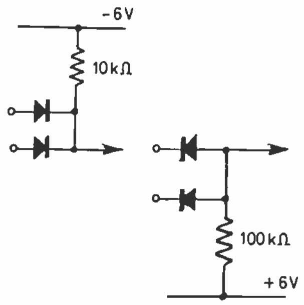


Fig. 1. Typical diode logic circuits were the AND and OR functions shown here. Their operation in this circuit depended on the OR gate following the AND gate since the current in the OR gate had to be a small part of that in the AND gate to maintain correct logic levels.

which followed brought Western Electric, of which Bell Laboratories were a part, eight million dollars in royalties in the first ten years of operation—and this was whilst industry was merely 'learning the ropes'.

At first all eyes were on linear applications such as communications. The earliest semiconductor digital circuits appeared in the years 1955–60. Amongst the first of these were diode logic gates[2] (Fig. 1). These were really scaled-down versions of thermionic circuits. Perhaps this was inevitable for any new circuit technology tends to lean heavily on its predecessor and some time generally elapses before a circuit design philosophy develops which is characteristic of the new technology.

Passive diode gates were not sufficient to form a versatile logic system since deterioration of voltage logic levels occurred through the system. It was therefore necessary to provide some level regeneration and simple single-transistor inverters were introduced to restore the voltage levels. Here again much of the original work was done at the Bell Laboratories[3] in about 1956 but the resulting diode transistor logic, or d.t.l., soon became commercially available. One form, marketed in UK by Mullard from about 1960, was called 'Combi-elements'. These modules were discrete circuits using germanium diodes and transistors, and later silicon diodes for OR gates. The circuits were encapsulated and could be wired directly to a printed circuit board. One of the problems of these systems was that the d.t.l. NAND or NOR gate was not at first seen to be a universal module.[4] Thus the range of Combi-elements included AND gates, OR gates, inverters and emitter followers to provide high fan-out driving capability.

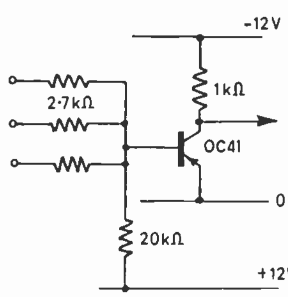


Fig. 2. The basic r.t.l. gate operated at a maximum speed of typically 100 kHz. The term 'r.t.l.' is preferred to the earlier[7,8] 't.r.l.' since the signal encounters the resistive network before the transistor.

The resistor-transistor logic circuit, r.t.l., was perhaps the first circuit to be seen as a universal module. In this circuit a resistive network replaced the diodes, and a transistor inverted the output of the network. The circuit design was based on analysis of the operation of the transistor as a switch. The theory of switching based on charge parameters was put forward at this time by Beaufoy and Sparkes[5,6] (Fig. 2). This form of the circuit was developed by a number of companies, notably by Westinghouse in 1956–57.[9] Commercial forms were available both in USA[10] and in UK in 1959. Two British forms were 'Minilog' developed by Elliott Automation, now a part of GEC, and 'Norbit' by Mullard. The Minilog circuits were made as 'log piles' that is, the components were mounted in cordwood. This consisted of two parallel cards carrying the connexions, between which the components were soldered. The complete circuit was later encapsulated. Minilog elements found application in a wide range of equipment including data loggers and process controllers such as the Elliott ARCH. Like the Norbit, germanium transistors were used, and due to the relatively good noise immunity compared with other systems then available, the modules found most application in industrial equipment where electrical noise was a problem. Depending on the logical convention employed, these gates were either NOR or NAND functions. Understandably the two companies chose different conventions. Hence a design philosophy was developed in which any function which could have been obtained with AND-OR logic could be realized with either solely NOR logic or NAND logic. Thus two levels of NAND logic are equivalent to an AND-OR system whilst two levels of NOR gates correspond to an OR-AND network. Further, the design of these gates lent itself to a statistically-based approach[11] using components of known tolerance, thus setting the scene for computer-aided circuit design later on.

These circuits were robust when encapsulated, were suitable for noisy environments and, with emitter followers, were capable of driving electro-mechanical devices. They appealed to those mechanical engineers who were suspicious of electronics and to whom their slow operating speed was often unimportant. The high-speed market was divided between a number of other competitors, one of which, resistor capacitor transistor logic, or r.c.t.l., merely by-passed the input resistors by 'speed-up' capacitors. Complementary p-n-p–n-p-n circuits[12] were also being designed. This was one area where valves could not be described as forerunners.

Another system, of more fundamental interest, was the current mode logic[13] or c.m.l., developed by IBM in 1956. Current mode logic gates can be fast when only a small voltage swing is permitted. However a reference voltage is necessary (see Fig. 3); also neither of the logic levels are at or near either of the supply potentials. In addition, practical circuits use substantial numbers of components and consume a relatively large power. They are though less dependent on the characteristics of the transistors than, for example, are d.t.l. gates. The noise immunity is small when the logical swing is restricted. All of this

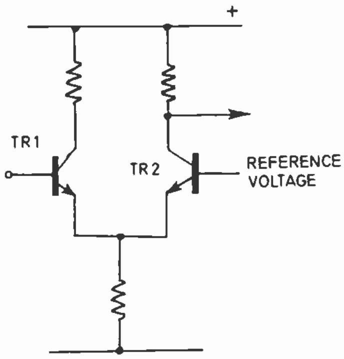


Fig. 3. A basic current mode logic gate has a reference voltage applied to the base of TR2 so that logic levels at the base of TR1 determine the levels at the output of TR2.

shows that the choice of logical system depended very much on the desired characteristics of the system and on those of the environment in which it was to be used, so that the decision was not always easy.

One other form of logical circuit was available during the period 1955–60: this was directly coupled transistor logic[14] or d.c.t.l. (Fig. 4), early work being done at Bell Laboratories, Raytheon and notably Philco. With d.c.t.l. the system was fast but the supply potential was low and logic levels were close so that noise immunity was a problem. Also when a number of bases were connected to one collector one of these would have a lower impedance than the others and so take most of the current. This 'current robbing' could be avoided by the use of resistors in series with each base. The circuit then resembled that of an r.t.l. gate.

The story of d.c.t.l. is closely linked with that of Philco who developed first barrier-layer transistors and later, in partnership with Plessey, as Semiconductors Ltd, microalloy-diffused transistors. Before 1939 Philco was the second largest radio manufacturer in USA. They attended the Bell Symposium in 1952 and three years later were one of the top three USA transistor manufacturers with over two-thirds of the American h.f. transistor market and second only to Bell in technology.[15] They believed in the future of digital circuits, principally d.c.t.l., for computers and at that time the transistors they developed were a considerable advance on the

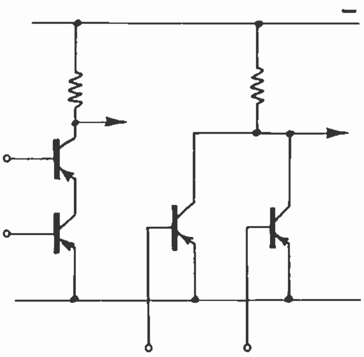


Fig. 4. D.c.t.l. NAND and NOR gates. The d.c.t.l. gate is only really practicable when the forward voltage at the base of a conducting transistor is higher than its collector saturation potential.

junction devices then available. Semiconductors Ltd was set up at Swindon in 1959 with Philco providing automatic equipment to manufacture transistors suitable for switching circuits. Plessey soon discovered that the automatic plant was only automated when controlled by graduates. Also the new planar process made Philco's transistors obsolete. They bought out the Philco interest and eventually the process was abandoned. Meanwhile Philco's computer division ran into trouble and production was discontinued. The sales and profits of Philco fell each year until it was bought by Ford in 1961. Philco-Ford finally closed in 1969 and d.c.t.l. had been lost somewhere along the road. The story of Philco is recounted here to illustrate the risks of bad management decisions or technological mistakes and may help to explain why company regrouping in UK has been necessary in an industry like digital electronics, to avoid such pitfalls here.

3 The Influence of Silicon

Germanium was an expensive raw material and in addition germanium transistors had a number of disadvantages when used as switches. Silicon was potentially plentiful, but at first the processing was expensive and silicon transistors were viable only for defence and

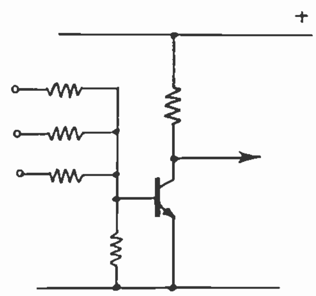


Fig. 5. The basic r.t.l. gate with a silicon transistor needed only one supply line due to the larger forward voltage necessary at its base to drive the transistor into conduction.

essential instrumentation. Texas Instruments[16] along with a number of other companies such as Ferranti commenced research with silicon in 1953 with varying success (Fig. 5). Texas produced their first silicon transistor in 1954 and had a virtual monopoly of the market for three years. Ferranti, for example, did not commence to sell transistors, albeit mesa devices, until 1958. Hence silicon logic circuits were slow in appearing[17] and when they did it was some years before they were cheap enough to achieve an extensive penetration of the general market. For instance a silicon version of Norbit, marketed as Norbit 2 with a maximum speed of 10 kHz, did not appear until 1968, by which time the planar process was well established and much of the market was looking elsewhere.

During this period digital instrumentation was beginning to become established and numerous articles published during this period evidence that both combinatorial and sequential logic were finding application not only in research but in commercial instruments such as frequency and phase meters,[18] and voltmeters.[19,20] By 1967, eight years after the planar process had been devised, R. K.

Richards[21] in a scholarly but not over-modern text, wrote that silicon was now unchallenged for digital circuits.

4 Integrated Circuits

It was the planar process that really made integrated circuits possible. Here again was a new technology and it took time for the appropriate circuit design philosophy to emerge. Active devices were no longer the most expensive, rather in terms of silicon area passive devices were often more expensive than transistors. Thus the design of gates had to be rethought so that capacitors were eliminated or minimized, resistors kept to a narrow range of values and the overall design based on relative tolerances rather than absolute ones.

At first digital integrated circuits were not monolithic and multi-chip circuits were the best that could be achieved, a situation which was to be repeated later with hybrid l.s.i. circuits. A typical example is the XK 75, a d.t.l. gate produced by ITT (then STC Ltd). This gate, more complex than the earlier d.t.l. circuits, is shown in Figs. 6 and 7. The silicon diffused epitaxial n-p-n transistors type BSY 28 and the diodes were bonded to a small printed circuit board mounted on the header. This d.t.l. gate had a total turn-on time of 14 ns and a turn-off time of 22 ns.

Other circuits of this period, produced by Marconi Microelectronics at Chelmsford used wire bonding

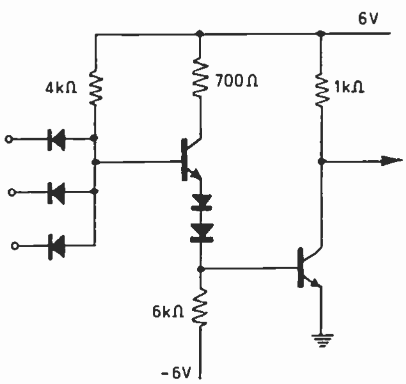
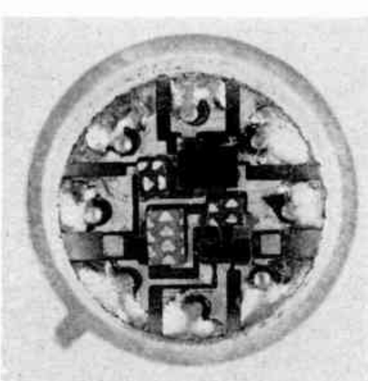


Fig. 6. STC multi-chip d.t.l. gate manufactured about 1965.



(Photo: E. Forbes)

Fig. 7. The circuit shown in Fig. 6 was mounted on a printed circuit card, held on an 8-pin TO5 header.

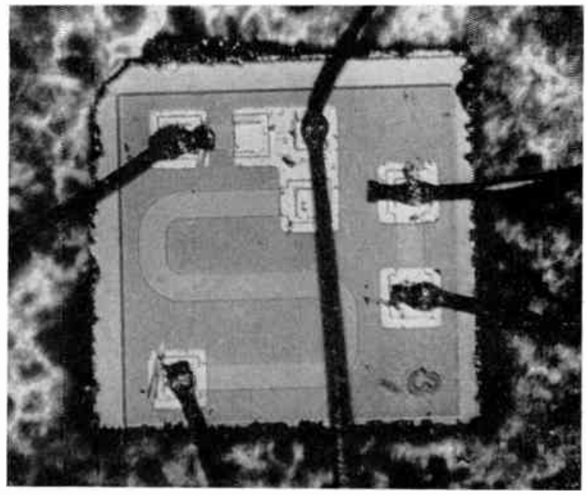


Fig. 8. A chip carrying a resistive network with wire interconnexions which was part of a multi-chip circuit.

between chips (Fig. 8). By about 1964 thick and thin film circuits[22] were being proposed using silicon chips. In 1965 Ferranti introduced Micronor II, a 9 ns integrated circuit. The use of d.t.l. circuits was increasing, a typical propagation time of 30 ns being quoted[23] for the general-purpose logic for the IBM 360 series.

Later silicon monolithic d.t.l. gates, such as the Fairchild 900 series of about 1967, manufactured by ITT and Marconi-Elliott amongst others, were somewhat similar to that in Fig. 6 except that a single power supply was used. They typically had a propagation time of 25 ns at a dissipation of 10 mW per gate with a fan-out of 8 and noise immunity of 1 V. From about 1966 plastic dual-in-line encapsulation had come into use[24], for d.t.l. as well as for other forms of logic. A few months later e.c.l. chips were being used for 16-bit read-only memories dissipating 270 mW in 14-lead plastic packages[25].

One of the main advantages of d.t.l. is that of direct wired connexion of a number of outputs from different gates. This facility known as 'wired-OR' can only be used so long as there is no possibility of the power supplies being short circuited by one gate being off whilst another is on. A wired-OR connexion is possible with other forms of logic, of course, but it was with d.t.l. that it made its greatest impact.

It is interesting to see the way in which the technology of monolithic integrated circuits influenced circuit design. For example, the basic r.t.l. gate of Fig. 5 was modified as shown in Fig. 9. This is because the surface area occupied by one transistor was similar to that for a 700 Ω resistor. Thus, a 5 kΩ resistor, if it could be avoided in circuit design, could be replaced profitably by a range of transistors, or by one or two transistors and a lower resistance value.

One of the arguments against current mode logic had been circuit complexity. This clearly was of less importance once solid circuits were used, particularly if active devices predominated. Also the reference supply needed could now either be on a single chip, as in the first Motorola m.e.c.l. series, or incorporated on to the logic

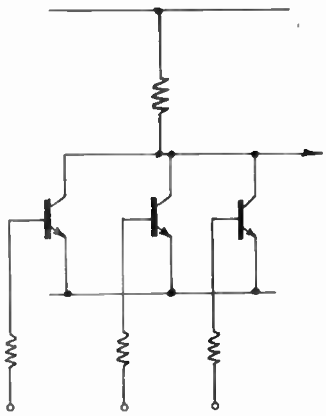


Fig. 9. An r.t.l. gate of the form used with silicon integrated circuits. Here two additional transistors had been incorporated to eliminate one resistor.

circuit chip. This is a logical system where the major impact has been made by Motorola[26] and by 1966 stage delays of less than 3 ns were being realized.[27] The higher speed, to-day less than 2 ns per gate,[28] has been achieved with a modification of the circuit (Fig. 10) known as emitter-coupled logic. This involves the addition of an emitter follower at the gate input which is repeated for each input limb. The input capacitance of the gate is reduced in this way, typically from about 7 pF to under 2 pF. This reduces the power-speed product so that e.c.l. can be designed to have only one third the dissipation of that of c.m.l. for the same operating speed. The logic levels for c.m.l., typically about 800 mW, which are the cause of the poor noise margins, can profitably be higher with e.c.l. One of the incidental problems of nomenclature in this area is that Motorola refers to c.m.l. circuits as m.e.c.l. and e.c.l. circuits as 'double level e.c.l.'. The latest versions of e.c.l. circuits are MECL 3 and Motorola 10 000 series.

Once the principles of solid state design had been appreciated, with its emphasis on a new balance between active and passive components, it was natural that a new type of logic gate should emerge which exploited this

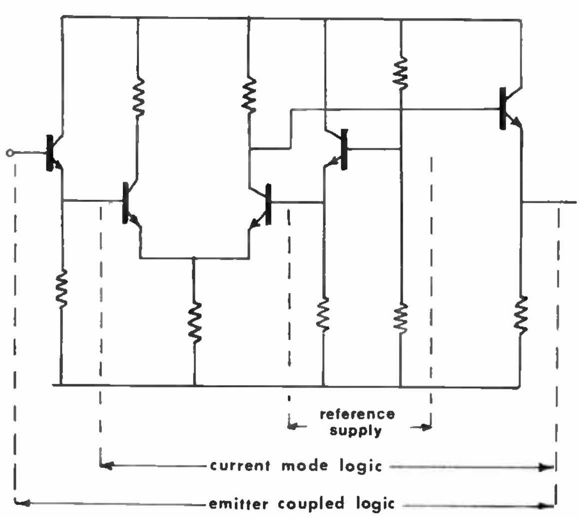


Fig. 10. Comparison of e.c.l. and c.m.l. circuitry.

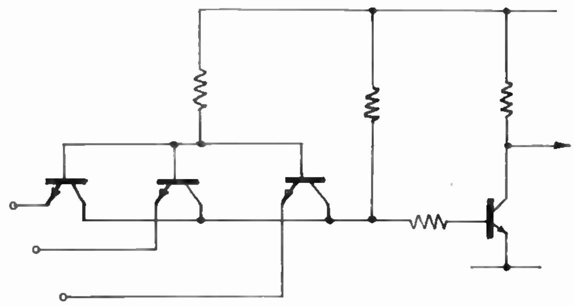


Fig. 11. The basic t.t.l. gate. Provided a sufficiently high value of output resistor was used it would be possible to connect a number of these gates together in a wired-OR configuration.

philosophy. This was transistor–transistor logic (Fig. 11). Early work on t.t.l. came chiefly from Fairchild[29,30] in 1962. Although potentially not as fast as e.c.l. this type of gate has achieved general acceptance in the 54 and 74 series developed by Texas and made under licence since 1967 perhaps more widely than any other logic system.

Amongst the developments in t.t.l. were the use of multi-emitter transistors, probably due to Plessey and the totem-pole output which gave greater fan-out, and improved the transfer characteristic and noise immunity, although with it a wired-OR connexion is no longer possible. Thirdly, there was the use of Schottky diodes[31], a Texas development. All these points are incorporated in the circuit in Fig. 12. The Schottky diode is an anti-saturating barrier diode formed by aluminium on silicon. It has a knee-voltage 0·2 V lower than for a silicon junction diode and is therefore able to prevent the transistor from saturating by providing an alternative current path. Whereas c.m.l. and e.c.l. often achieve speed by not running into saturation, other systems with larger swings use transistors which are driven heavily into conduction. Hence they have to rely on gold to dope the silicon to reduce charge storage time. Gold doping of chips minimizes carrier lifetime, so that the chips therefore have characteristics which make them unsuitable for linear amplifiers. When high speed is the result of Schottky diodes rather than gold doping, digital and linear circuits become compatible on the same chip.

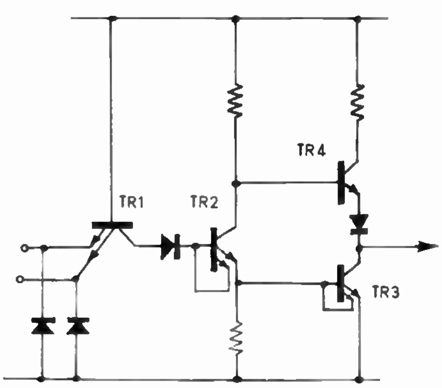


Fig. 12. A t.t.l. gate with a multi-emitter input transistor, TR1, two transistors with Schottky diodes, TR2 and TR3, and a totem-pole output TR3 and TR4.

5 Other Forms of Logic

In this brief account of the development of circuit design for a semiconductor industry in which the technology was itself in a state of continual development it is not possible to deal at length with all of even the important variants on design such as the use of tunnel diodes[32] for logic and counting circuits at frequencies up to 1 GHz which seemed to be so attractive about 4 years ago. Also there were 2-phase (Fig. 13) and 4-phase dynamic bipolar circuits which relied on the charge stored in the collector-substrate capacitance and found application as random access memories with cycle times of the order of 150 ns. However, two developments more perhaps than any others have made interactive contributions in this area.

The first of these was the JK flip-flop[33] suggested by Texas engineers in 1964. Flip-flops or binary stages are the fundamental memory elements of semiconductor shift registers, counters and stores. The set/reset flip-flops of that time had the disadvantage that under certain input conditions it was not possible to predict the future state of the flip-flop. The JK flip-flop had routing gates around it to avoid this indeterminate condition.

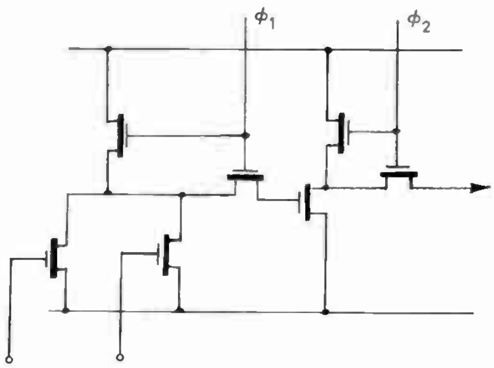


Fig. 13. Two stages of two-phase dynamic m.o.s. circuitry having a typical dissipation of 3 mW per stage. The circuit operates by transient capacitor storage. M.o.s. devices which are part of such a circuit can be at least an order smaller than comparable bipolar ones.

It is interesting to note that early JK flip-flops had typically 14 transistors on the chip with as many resistors. A year or so earlier it would not have been possible to put 28 components on a chip less than 0·9 mm square with any hope of an economic yield and it would certainly not have been economic to construct it with discrete components. Thus the JK flip-flop is an example of a circuit design which could only be exploited because of the advanced state of silicon technology.

The second development of note was the introduction of the m.o.s. process to make unipolar logic gates.[34] Discrete m.o.s. logic elements were first made commercially available in about 1964 by General Microelectronics, later taken over by Ford-Philco, but the technological surface problems of passivation and catastrophic failure due to electrostatic pick-up were not solved until about 1968. Although slower than bipolar circuits the concept of 'all active devices' on the chip where the physical size of devices was related to circuit impedances was very attractive to designers. The significance of m.o.s. circuit design was that it represented a further step in the philosophy that solid-state design favoured active devices at the expense of passive ones. The chief drawback of the system was that it was essentially a high impedance one and not compatible with, say t.t.l. gates, without modification.

M.o.s. circuitry flourished in integrated form, barely even pausing at the stage of simple gates, to develop into dynamic and static shift registers, various types of memory systems and, by 1972, complete arithmetic units on a single chip. With bipolar circuits this progression was much slower and the final end point much less obvious.

6 Larger Digital Systems

In the last six years there has been enormous progress towards increasing the number of components on commercially available chips, progress which has been accompanied by greatly improved yields in processing. The reasons for this have been as much political as technical with large, principally US, government contracts connected with defence and aerospace which funded the research and production facilities. Digital circuit and sub-system design took two lines of development.

The first of these was the hybrid approach in which silicon chips were bonded to a thick film pattern on a ceramic substrate. This is still a 'bread and butter' line for many companies. The chips could be individually tested and the yield on the hybrid system could be expected to be high. An extreme example of this technique in which the chips were mounted on substrates and then stacked was the Hamilton standard microcircuit module. Devised by the United Aircraft Corporation and introduced into UK by Hawker Siddeley Dynamics in 1967 as the HSD Micropack it consisted of 12 alumina substrates about 1 cm (0·4 in) square in the stack. These were interconnected at the edges so that each module could contain a sizeable sub-system.

It was claimed that they were an order more reliable than conventional circuits. Not a great improvement! In fact reliability was the keynote to much of this work. Digital circuit designers who had found the transistor to be small were not always able to take advantage of the further reduction in size which integrated circuits brought. It was observed that equipment sizes did not shrink with these innovations: the contents merely became more complex. Circuitry becomes more sophisticated, more versatile and so tends to fill the space available for it. The real test is whether the reliability of the equipment has risen at all. Also as circuits become more complex there is a sharp rise in the cost of testing them. Really the hybrid circuits and the stacks were repetitions of the earlier *ad hoc* multi-chip philosophy and where reliability has not been acceptable they have been discarded.

The second line of development was to put still more on the chip. Medium-scale and large-scale integration were terms used to illustrate this. Thus there has been a great increase in the number and complexity of digital functions now available. These include gated full adders, programmable shift registers and binary-decimal converters. We saw at one time complete chips probe-tested

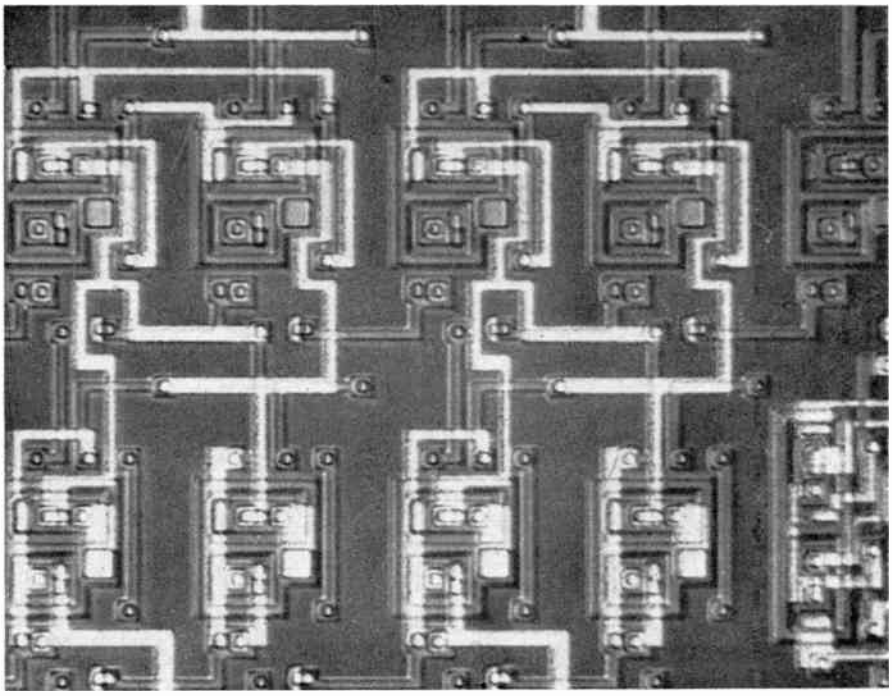


Fig. 14. A section of the circuitry of the Ferranti uncommitted logic array in which the complete array is held on a chip 2 mm × 3 mm. This is typical of advanced techniques in 1973, being produced using c.d.i. technology.

(Photo: E. Forbes)

and given discretionary wiring[35]; surely the monolithic equivalent of the multi-chip philosophy (Fig. 15). Now new, simpler, and so more reliable, processes[36] are being used, such as Fairchild's Isoplanar and Ferranti's collector diffused isolation (c.d.i.)[37] so that many hundreds or even thousands of gates can be put on one chip.

7 Arrays

One feature of the design of digital circuits which has been a continuing theme over the last nine years is that of having a solution (large scale integration) in search of a problem. What is there that is sufficiently complex to warrant the use of l.s.i. and at the same time is so general and hence has so much sales potential, that production costs can be kept low? The recurrent answer has been that of a general-purpose logic array. In 1964 it was the Texas series 53 Master Slice with about 70 components on a 7·65 × 3·8 mm (0·065 × 0/150 in) silicon chip which the manufacturer probe-tested and hence selected for metallizing. Hardly l.s.i., but a step in that direction. In 1965 Motorola[38] announced their 'Polycell approach to l.s.i.'. This used discretionary wiring so that faulty areas could be probe-tested and excluded from the metallization pattern. At that time a typical example quoted was a 16-bit content addressable memory. However it was admitted that this had only a development status and was not in production. Some years later though, with improvements in yield, discretionary wiring has been all but forgotten.

In 1966 Fairchild were developing their Micromatrix chip. This was an array of 4 × 5 cells each consisting of 4 m.o.s. resistors and 12 m.o.s. transistors interconnected to form NAND gates. The propagation rate was 40–60 ns per gate and when operating at 1 MHz their dissipation was 1 mW per gate. In this case the manufacturers produced chips with each cell uncommitted to any special function. The buyer was to specify to the manufacturer the nature of the function so that a further metallization layer could be used to produce custom-made chips.

In 1972 Ferranti introduced the uncommitted logic array (Fig. 14), based at first on r.t.l. gates modified by Schottky base-emitter diodes, and later by d.c.t.l. gates of a similar kind, with however, base input resistors. This u.l.a. was thus a bipolar system based on collector diffusion isolation in which heavily-doped collector contacts in the silicon chip effectively isolated each transistor without the necessity for lands—so economizing on surface area.

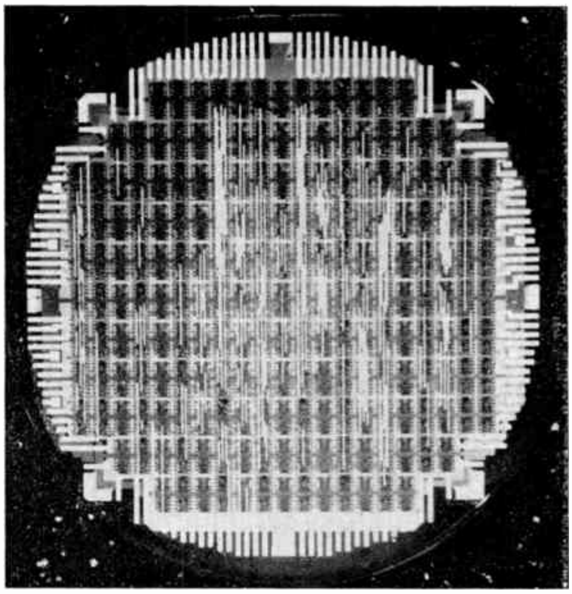


Fig. 15. An example of discretionary wiring typical of advanced techniques in 1968, produced by Texas Instruments Ltd.

Thus the idea of a general-purpose logic array has been around for some time: not however that it has grown ever larger. The l.s.i. being proposed in 1968 were single 1-inch slices, each slice being devoted to one complete system and having typically 100 leads (Fig. 15). An essential feature was multi-layer metallization. In 1973 less emphasis is laid on multiple metallizations even though the flatter surfaces of processes like c.d.i. might make this attractive. There is also less emphasis on complete slice arrays since the surface area now needed for a complex function, such as a logic array or a semiconductor store is less than in 1968.

To some extent l.s.i. is a self-regulating design philosophy with feedback provided by the initial dilemma of a solution seeking a problem. It has not been the purpose of this paper to discuss the problems of yield, but clearly here too there is feedback for the greater the number of processes and the greater the surface area the less the yield is likely to be.

To-day we view the prospect of digital systems in communications, control, computation and instrumentation which are highly reliable, and relative to those of 25 years ago, extremely complex. Circuit design has followed closely on changes in technology but due to the nature of design to-day and to the changes in the industrial scene which have accompanied the technological changes, the circuit designer is a much more skilled but less often met type of engineer than formerly. In time he may be the only exponent of the art of design in electronics engineering as we have known it over this last quarter century.

8 Bibliography

1. Caldwell, S. H., 'Switching Circuits and Logical Design', (Wiley, New York, 1958).
2. Strong, P. F., 'Rectifiers as elements of switching circuits', *Proc. Assn. Computing Machinery*, May 1952.
3. Cagle, W. B. and Chen, W. H., 'A new method of designing low-level high-speed semiconductor logic circuits', *IRE Wescon Convention Record*, **1**, Pt. 2, pp. 3–9, August 1957.
4. Young, J. F., 'Transistor-diode static switching units', *Electronic Engineering*, **34**, pp. 595–9, September 1962.
5. Beaufoy, R., 'Transistor switching circuit design using charge parameters', *Proc. IEE*, **106**, p. 1085, 1959.
6. Beaufoy, R. and Sparkes, J., 'The junction transistor as a charge controlled device,' *ATE Journal*, **13**, No. 4, p. 310, 1957.
7. Dean, K. J., 'An Introduction to Counting Techniques and Transistor Circuit Logic', (Chapman and Hall, London 1964).
8. Wray, W. J., 'Worst case design of variable threshold t.r.l. circuits', *IRE Trans. on Computers*, **EC-11**, pp. 382–90, June 1962.
9. Rowe, W. D. and Royer, G. H., 'Transistor NOR circuit design', *Communications and Electronics*, **31**, pp. 263–7, July 1957.
10. Simkins, Q. W., 'Transistor-resistor logic', *Semiconductor Products*, pp. 34–8, April 1959.
11. Marcovitz, M. W. and Seif, E., 'Analytical design of resistor-coupled transistor logical circuits', *IRE Trans. on Computers*, **EC-7**, pp. 109–19, June 1968.
12. Rubinoff, M., 'A new family of transistor switching circuits', *Communications and Electronics*, **31**, pp. 286–9, July 1957.
13. Henle, R. A., 'High speed transistor circuit design', *Proc. Eastern JCC*, **10**, pp. 64–6, December 1956.
14. Harris, J. R., 'Directly coupled transistor logic circuitry', *IRE Trans. on Computers*, **EC-7**, pp. 2–6, March 1958.
15. 'Gaps in Technology—Electronic Components', OECD, Paris, 1968.
16. Haggerty, P., 'Strategies, tactics and research', *Research Management*, p. 9, June 1967.
17. Harris, J. N., Gray, P. E. and Searle, C. L., 'Digital Transistor Circuits', (Wiley, New York, 1966).
18. Fuchs, H. and Wheable, D., 'An all-electronic 4-digit digital voltmeter', *J. Brit. IRE*, **20**, p. 541, 1960.
19. Rose, J. A., 'Digital *Q* meter counts damped oscillations', *Electrical Design News*, p. 37, August 1963.
20. Bell, E. C. and Leedham, R. V., 'Digital measurement of mean phase shift', *Electronic Engineering*, **35**, p. 664, August 1963.
21. Richards, R. K., 'Electronic Digital Components and Circuits', (Van Nostrand, New York, 1967).
22. Fuschillo, N. and Kroboth, J., 'Multifunction monolithic thin-film compatible d.t.l. logic circuits for data processing', *IEEE Trans. on Computers*, **EC-15**, pp. 640–6, August 1966.
23. Hahn, E. F. and Tomczak, J. J., 'General purpose 30 ns logic family for IBM system 360', *Electromechanical Design*, **10**, pp. 38–43, August 1966.
24. Nixen, D. 'Packaging, testing and reliability', Paper 11, Microelectronics in Equipment Conf. London, November 1967.
25. Murray, D. E., 'New developments in semiconductor memories', Paper 26, Proc. Microelectronics in Equipment Conf. London, November 1967.
26. Roehr, W. D., 'Techniques of current-mode logic switching', *Electronic Design*, **10**, pp. 54–9, September 1962.
27. Crowther, G. O., 'E^2CL—First integrated logic element', Paper C.1, Proc. Conf. Microelectronics in Equipment, London, November 1966.
28. Kinniment, D. J. and Edwards, D. B. G., 'Circuit technology in a large computer system', Conf. Computers—Systems and Technology, London, October 1972. (IERE Conference Proceedings No. 25.)
29. Beeson, R. H. and Ruegg, H. W., 'New forms of all-transistor logic', *Proc. International Solid State Circuits Conf.*, **5**, pp. 10–11, 104, February 1962.
30. Ruegg, H. W., 'Transistor–transistor logic circuits', *Electronics*, **36**, pp. 54–7, March 1963.
31. US Patent 3463975, filed 1964, issued 1969.
32. Charters, J. S. T., 'The design of logic and counting circuits capable of operating at 1 GHz pulse repetition rate', Conf. Digital Methods of Measurement, Canterbury, July 1969, pp. 39–58. (IERE Conference Proceedings No. 15.)
33. Cook, C. R. and Martin, B. M., 'New semiconductor networks reduce system complexity', *Electronics*, **37**, pp. 25–9, January 1964.
34. English, P., 'Digital applications of m.o.s. integrated circuits', *Marconi Review*, **31**, pp. 171–88, third quarter, 1968.
35. Dean, K. J., 'The development of large scale integration', *The Radio and Electronic Engineer*, **37**, pp. 135–40, March 1969.
36. Roberts, D. H., 'A Comparison of High Density Bipolar Technologies', Allen Clark Research Laboratories, March 1972.
37. Murphy, B. T. *et al.*, 'Collector diffusion isolation integrated circuits', *Proc. IEEE*, **57**, 1523–7, September 1969.
38. Hazlett, L., 'The polycell approach to large scale integration', Motorola Development Report, 1965.

Manuscript received by the Institution on 27th November 1972. (*Paper No. 1497/CC 161.*)

UDC 621.382.029.65

Semiconductors for microwave frequencies

Professor H. HARTNAGEL,
D.Eng., C.Eng., F.I.E.R.E., M.I.E.E.*

SUMMARY

Microwaves can now be processed by semiconductor structures, which are either a further development of the transistor for the lower microwave frequency ranges, or which use the non-linear behaviour of junction devices for mixing and parametric effects, particularly for low noise levels, or which exploit transit-time effects to create an effective negative conductance g_d such as impatt and Gunn devices. The important rôle of material parameters is discussed New experimental results on g_d vs. microwave voltage amplitude are presented for Gunn diodes as an illustration of the type of device characterization required.

* *Department of Electrical and Electronic Engineering, University of Newcastle upon Tyne NE1 7RU*

1 Introduction

In the same way as concerted research efforts led to a replacement of the triode valve by the transistor with its advantage of small size, long life and often cheap mass production, impressive progress has been made in substituting semiconductor devices for microwave tubes. This has only been partly possible by extending the principles of transistors into the microwave range, because transit-time effects limit the operating frequencies. The entire microwave frequency range can only be covered by new device ideas, where transit-times are especially made use of to obtain signal gain. This was of course originally the same situation with the development of microwave valves such as klystrons where transit time limitations were overcome by incorporating charge-carrier transits for the amplification process. A corresponding semiconductor example is the impatt diode, where a bunch of charge carriers drifts across a drift length for almost a whole period of the microwave signal so that the device current remains high even when the microwave voltage takes on a negative value. A high current for a negative microwave voltage represents an effective negative conductance at the frequency of operation.

Additionally, new semiconductor phenomena are available for generating and amplifying microwaves where no direct microwave-tube equivalence can be quoted. Here one has the transferred-electron effect, which results from the electron behaviour in some semiconducting crystals.

It is of course essential to investigate the microwave properties of semiconducting materials systematically and an outline of the relevant features is presented in the next section. The subsequent sections deal with microwave transistors. Then mixers and parametric amplifiers with semiconductor devices are discussed. Other junction devices suitable for microwave applications are the tunnel diode and the avalanche devices, which are treated in Sect. 6. Finally transferred-electron active components are presented.

2 Semiconducting Materials for Microwaves

At microwave frequencies the semiconducting properties of the materials used can cause a skin-effect, which is defined by the depth δ of the penetration of microwave currents into a conducting medium. Maxwell's equations give here:

$$\delta = \left(\frac{2}{\omega\mu_M\sigma}\right)^{\frac{1}{2}}$$

where ω = angular frequency, μ_M = permeability of the medium, and σ = conductivity. This means, for example, that for a charge carrier density of about $10^{16}\,\text{cm}^{-3}$, corresponding to conductivities of around $10\,\Omega^{-1}\,\text{cm}^{-1}$ for good semiconductors, the skin-effect alters the microwave current density if the device dimensions become larger than $2\mu m$ for Q-band frequencies.

The microwave behaviour of semiconductors is often strongly affected by the life times τ_L of excess carriers. These occur with bipolar junctions, where majority carriers from one side of the junction diffuse into the opposite side and exist there for a long time as minority carriers. When one wants to switch such a junction from the forward to the reverse state, the charge represented by these excess carriers has first to be removed. This can take many nanoseconds and represents usually serious limitations regarding the operation at microwave frequencies, although there are some possibilities of making special use of the resulting current wave shape for microwave processing (for example, when using the step-recovery diode). Excess carrier life times are also relevant where carrier generation by either photons, high fields or electron impact occurs. This can be of disadvantage as, for example, with the light-detector diode for high-bit-rate optical pulse modulation, where it is difficult to detect light pulses of much less than one nanosecond duration due to excess-carrier life times. Of course, by proper use of excess carrier life times, new active microwave structures are possible such as the one proposed by Read in 1958.[1]

There are several further semiconductor time-constants which sometimes have relevance for microwave operation. The momentum relaxation time τ_m, gives the average time between collisions of charge carriers with phonons, impurities, dislocations or other crystal imperfections. This constant is involved with the definition of the carrier mobility μ, i.e.

$$\mu = \frac{|e|\tau_m}{m^*}$$

where e is the electronic charge and m^* is the effective mass of the freely mobile charge carrier. The momentum balance equation is here

$$eE\tau_m = m^*v$$

where E is the electric field and v the drift velocity. In an analogous manner, the energy relaxation time τ_e is defined by the energy balance equation, namely

$$\varepsilon - \varepsilon_0 = eEv\tau_e$$

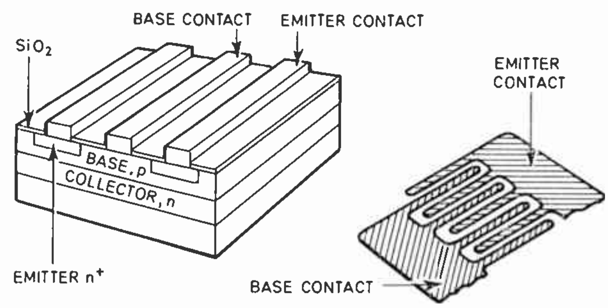
where $\varepsilon - \varepsilon_0$ is the excess energy of the electrons over their thermal equilibrium energy ε_0. Whereas τ_m can be measured easily, it is difficult to determine τ_e, although its existence has a pronounced effect on the so-called hot-electron regime, which occurs for high electric fields where velocity saturation, transferred-electron effects and other phenomena occur which are usually relevant for microwave device structures.

The advantage of Si is that its device fabrication technology is mastered to a very high degree. This is often accepted as a convincing reason for using mainly Si rather than the many other semiconductors available. However some of the other semiconductors have additional properties which cannot be found with Si. Firstly, for microwave transistors there is the question of the higher room temperature mobility of several compound semiconductors. The highest is that of InSb, whose energy gap is unfortunately extremely small (0·16 eV) so that this material cannot be used for the development of junction devices or space-charge layer f.e.t.s. Avalanching would occur too rapidly. However, GaAs, which has quite a large energy gap (namely 1·34 eV, whereas Si has 1·1 eV), exhibits also an increased mobility (namely 8000 cm^2/Vs, against 1600 cm^2/Vs for Si). Therefore GaAs f.e.t.s have succeeded in reaching the highest frequency ranges attained by transistors. Here an important parameter is also the magnitude of the saturation velocity, as this structure is of course operated with as high a velocity as possible in order to reduce transit times. It seems that electrons in GaAs have a drift velocity which is about double the value of those in Si, although the electron-transfer effect can reduce the saturation velocity for increased fields again so that the advantage diminishes.

There can be other useful properties with various non-Si materials. The electron transfer effect in GaAs, InP and some other semiconductors produces a negative differential mobility. Future device developments might be possible with semiconductors exhibiting ferroelectric, piezoelectric or ferromagnetic properties.

3 Bipolar Transistors

Bipolar microwave transistors are generally silicon planar epitaxial-diffused n-p-n structures. The emitter is designed to increase the active emitter-base junction area. At the high operating currents the current across this junction is pushed out towards the edges so that the active area is strongly reduced. One therefore increases the emitter edge length by various types of interdigital structures. A simple example is shown by Fig. 1, where the emitter and base are built like a set of interlocking combs. With photoetching techniques, the strip width can be as low as 1 µm or less, particularly if, instead of optical exposure, electron beam processing of the resist is performed.



(a) Sectional view. (b) Top view of interdigital structure.

Fig. 1. An interdigital transistor

A further development is the overlay structure of Fig. 2, where many small separate emitter sites are used instead of the continuous emitter strip of the interdigital geometry. This provides a substantial increase in emitter periphery without requiring an increase in physical area of the device. The fabrication process involves an additional diffused region in the base to serve as conducting paths. This p^+ region offers three advantages: (i) it distributes the base current uniformly over all the separate emitter sites, (ii) it reduces the distance between emitter and base, and (iii) it reduces the base resistance and contact resistance between the metallization and the semiconductor. The term 'overlay' is derived from the fact that the emitter metallization lies over the base instead of adjacent to it. Of course the two types of area are insulated from each other by a SiO_2 layer.

An important limitation of the operating frequency is given by the charging time T_e of the emitter capacitance C_e by the emitter current. One can say approximately that $T_e = kTC_e/eI_e$ where I_e is the emitter current and $kT/_e = 26$ mV at 300 K. C_e is proportional to the total emitter-base junction area. As I_e is given by the length of the periphery, T_e can be kept small by a large ratio of emitter edge to emitter area. The overlay geometry is here advantageous too.

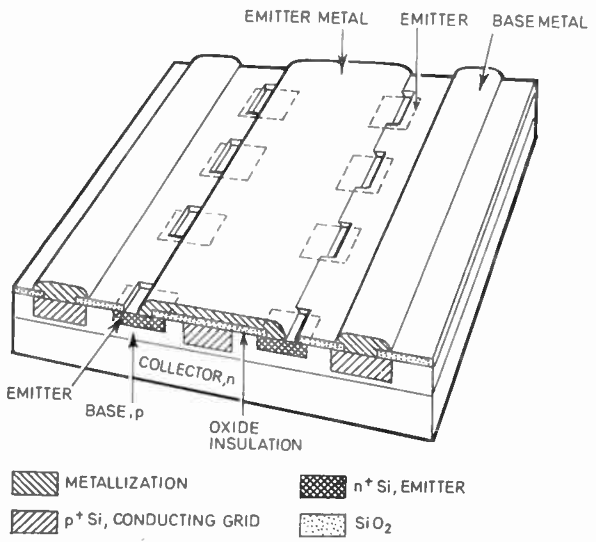


Fig. 2. An overlay transistor.

E. O. Johnson[2] showed that the ultimate performance limit of a transistor is set by the equation

$$V_m f_t \simeq E v_s / 2\pi \tag{1}$$

V_m is the minimum allowable applied voltage, f_t the cut-off frequency, E the dielectric breakdown field of the semiconductor and v_s the saturated drift velocity of the charge carriers. This equation assumes that the ultimate limitation of transistors is given by the shortest time it would take electrons to travel across a distance in the semiconductor, where V_m cannot be reduced below some reasonable value as otherwise noise fluctuations and similar considerations would take over the operation of the device. This equation is therefore equally valid for both field-effect transistors and for bipolar transistors if the charge carrier transfer across the base and the emitter capacitance C_e is taken into account. Often, the base current is not given by diffusion as is usual with transistors, but is due to drift with a high electric field. Microwave transistors have then a doping gradient across the base such that a suitable electric field is established across it. This field can be sufficiently high to effect charge carrier transfer with saturation velocities.

As E and v_s of equation (1) are known material constants, the transistor-frequency limitation for various semiconductors can be estimated. In fact, for the most important semiconductors the expression $Ev_s/2\pi$ is as follows:

GaAs : 1×10^{12} V s^{-1}
Si : 2×10^{11} V s^{-1}
Ge : 1×10^{11} V s^{-1}

Recent developments of transistors have come rather close to these limits.

The bipolar transistor has an overall time-constant which involves two further time-constants, the collector space charge transit time T_{sc} and the collector charging time T_c. The former can be approximated by the collector space charge width w_c and the carrier velocity v, i.e.

$$T_{sc} \simeq \frac{w_c}{2v}. \tag{2}$$

The latter is given by the collector-base capacitance C_c and the collector series resistance r_c, namely

$$T_c = C_c r_c. \tag{3}$$

Taking the time-constant due to base transit as T_b, the values for a typical 6 GHz Si bipolar transistor are as follows: $T_e = 10$ ps, $T_b = 2$ ps, $T_{sc} = 12$ ps and $T_c = 1$ ps. Clearly T_e and T_{sc} are limiting parameters.

It is interesting to note that both Si bipolar and narrow-gate Si field-effect transistors with $f_t \simeq 10$ GHz have been produced with $V_m f_t$ products above 100 V GHz, which is rather close to the theoretical limit of 200 V GHz. Further extensions in frequency will be obtained by sacrificing voltage capability. This means for bipolar transistors, further controlling the collector space-charge width by increasing the collector doping density.

The present status of microwave bipolar-transistor oscillators gives typically 1 or 2 W c.w. at 4 GHz. The best performance at present shows 5 W c.w. at 4 GHz (efficiency 30%, transistor power gain 4 dB).[17] The power-conversion collector efficiency is impressively high; for example at 2·2 GHz, microwave transistors give 7 W with 6 dB gain and 40% collector efficiency. This collector efficiency is defined as the ratio of maximum microwave power output at the frequency of interest to the d.c. power input. The efficiency is strongly dependent on the equivalent output admittance and on the heat dissipation capabilities. Very important are package considerations because package reactances can considerably affect transistor performance at microwave frequencies. Microwave transistors are best characterized by scattering parameters S, which can easily be measured by a microwave network analyser. These parameters are defined by the microwave voltage E of the incident (subscript i) and reflected (subscript r) waves for the input (subscript 1) and output (subscript 2) terminals, namely:

$$E_{r1} = S_{11} E_{i1} + S_{12} E_{i2}$$
$$E_{r2} = S_{21} E_{i1} + S_{22} E_{i2}$$

The maximum gain for matched input and output conditions is then

$$G_{max} = \frac{|S_{21}|^2}{(1-|S_{11}|^2)(1-|S_{22}|^2)}.$$

An important characteristic of a small-signal microwave transistor is the minimum noise factor, which is determined predominantly by the base resistance and the cut-off frequency. An improved structure has recently been developed[7] by employing lateral diffusion effects and ion implantation so that the noise figure was reduced to 2·3 dB at 4 GHz and 4 dB at 8 GHz.

The progress made with bipolar transistors is impressive, but it cannot be expected on the basis of the limiting time-constants discussed above, that they will compete above 10 GHz with the new types of microwave devices now avilable, which avoid the difficulties of transit time effects.

4 Field-Effect Transistors

The field-effect transistor does not have to rely on diffusion processes if Schottky gates are employed and can therefore be fabricated more easily. A Schottky junction has the additional advantage that it can be switched from the forward into the reverse direction without the need first to discharge the minority carrier charge. This makes Schottky-gate f.e.t.s particularly suited for fast switching applications. So far, diffusion processes in GaAs have not been very successful so that no competitive GaAs bipolar transistor has been developed yet. This is different from Schottky-gate f.e.t.s, where most impressive results have been achieved with GaAs, because the saturated drift velocities are larger than those for Si. According to equation (1), f_t values up to 30 GHz can be expected and practical gain values have been achieved at 15 GHz with laboratory samples.[3] The device was made on a 0·3 Ω cm, 0·5 μm thick n-type GaAs crystal grown epitaxially on a semi-insulating substrate with a gate width of 1 μm, and a channel length of 2 μm. A four-stage amplifier showed 16 dB power gain and 3 dB bandwidth of 150 MHz. At 10 GHz, c.w. output of 4 mW was reported. The noise figure of GaAs Schottky-gate f.e.t.s are 5 dB at 8 GHz and 10 dB at 12 GHz.

Again, f.e.t.s are transit-time limited and correspondingly the frequency of operation is fundamentally limited.

5 Mixers and Parametric Amplifiers

An early approach to cover the microwave range by semiconductor devices was made successfully by employing non-linear components. Applying power from a v.h.f. transistor generator to a non-linear multiplier diode of suitable characteristics enabled one to produce microwave power at some suitable harmonic. Similarly by applying a microwave signal together with a local-oscillator microwave to a non-linear element, the intermediate frequency could be obtained which could then be amplified by a v.h.f. transistor circuit before being upconverted again to microwave frequency ranges. Typical conversion losses of commercial mixer diodes above X band are 4 dB at 9 GHz, 5 dB at 20 GHz and 5·5 dB at 45 GHz. These figures are, however, highly dependent on the circuit configuration which has to be carefully designed and constructed. To maintain a low noise figure for a mixer–i.f. amplifier, the noise performance of the i.f. amplifier has to be optimal. This can mean that the intermediate frequency is best chosen to be below 0·8 GHz. This can of course limit the band width of the mixer–i.f. amplifier.

Transistors can be used to operate as amplifier-multipliers and oscillator-multipliers to produce power output at frequencies well above their maximum fundamental frequency of oscillation. The non-linear element primarily responsible for harmonic generation in transistors is the depletion-layer capacitance of the base-collector junction. This junction acts similarly to that of a variable-capacitance diode, the varactor diode. With an input of 1 W at 500 MHz, an output of 1·8 W at 1·5 GHz was obtained some time ago; the collector efficiency was then 20%.[4]

Varactor diodes have today been perfected for optimum frequency multiplication. Typically, the following c.w. output power can be obtained by transistor-driven varactor harmonic generation chains: 5 W at 8 GHz and 1 W at 13 GHz. Here several transistors operate in parallel on parallel chains of varactor multipliers. For example with an input frequency of 4 GHz an output of 1·2 W at 12 GHz was obtained with a conversion efficiency by a varactor-diode system of 62%.[5]

For multiplication ratios greater than 4, step-recovery diodes are generally used. These are varactor diodes which are specially designed to utilize the charge storage by minority carriers so that the effective non-linearity of the device is greatly improved. Typically for 5 W input at 2 GHz an output of 350 mW can be obtained at 10 GHz.

In those cases where the noise figure of an amplifier has to be kept very low a parametric amplifier can be employed. So far this circuit has been used primarily at cryogenic temperatures and extremely low noise figures can be achieved. Recently[6] attractive noise performance was obtained with uncooled parametric amplifiers.

Parametric amplifiers use the voltage dependence of the junction capacitance of varactor diodes. Manley and Rowe developed the relevant formulas, which show the relation between the power flow at various frequencies to and from the circuit. Usually, a signal frequency f_1 together with the frequency f_0 of a source of motive power to the amplifier is applied to the circuit incorporating a varactor diode; f_0 is termed the pump frequency. The difference frequency, $f_0 - f_1$, which is essential to the operation of the device, is called the idler frequency. Theory shows then that if the pump power W_0 is applied to the circuit, both signal and idler power are generated by it. If W_0 is not too large, the signal power applied will be amplified regeneratively. The system operates as a reflexion amplifier. The amount of power involved at each frequency is proportional to the frequency. The idler power has to be accommodated properly, otherwise no output signal is obtained. The internal noise in a completely lossless parametric amplifier originates only in the idler load. If the signal frequency is lower than the idler frequency, a very much reduced noise figure is possible.

Regarding the constructional details of microwave junction diodes one has either p-n, p-i-n or Schottky barrier diodes. These three types of components can be made to operate in various ways. Variable resistance diodes (often called varistors) are useful for down-converter mixers and fast rectifiers. Variable reactance diodes (varactors) are employed for up-converter mixers and parametric amplification.

6 Other Microwave Junction Devices

For very steep doping gradients, the conduction band of the n-type region is opposite the valence band of the p-type area with a narrow barrier between them. For small forward voltages, a tunnelling current occurs before an appreciable forward current sets in. Correspondingly, a voltage-current characteristic is obtained as shown by Fig. 3. A negative-resistance region occurs in the range from about 100 to 350 mV for Ge devices. The tunnelling current is no longer caused by drift or diffusion of carriers. As the tunnelling process occurs with the velocity of light, there are no transit-time frequency limitations.

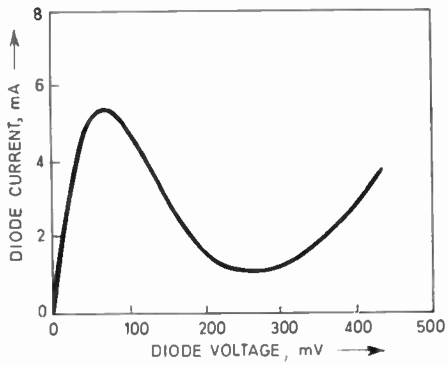


Fig. 3. Current–voltage characteristic of a tunnel diode.

The capacitance of a typical Ge tunnel diode in the negative resistance region is of the order of 5 μF/cm². This means that in actual diodes a capacitance of 50 pF is usually accompanied by a negative differential resistance of 1 Ω. The spreading resistance is small, because of the heavy doping of the diode. The upper frequency limit is given when the resistive part of the total diode impedance becomes positive. The lead resistance and various other parts are responsible for a lossy series resistance R_s so that the equivalent circuit of Fig. 4 can be found. The total impedance has then its resistive part changing sign at the frequency

$$\omega = \frac{\sqrt{R-R_s}}{\sqrt{R_s}RC}.$$

As the tunnel diode impedance is rather low it is difficult to match these devices to the normal 50 Ω transmission lines. Another disadvantage is the fact that the negative resistance goes down to zero frequency so that problems with parasitic oscillations often arise. Tunnel diodes have found common use in preamplifier stages because of the relatively low noise figures achievable. For example, they have been employed in satellite communication systems, where, however, the low saturation power (−15 dBm), difficulties in stabilization, and the reliability problem have been considered unfortunate drawbacks.

Ideas of directly incorporating transit-time effects in

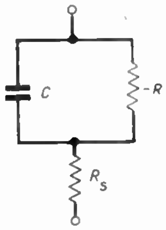


Fig. 4. Equivalent circuit of tunnel diode.

the production of an effective negative conductance for generation and amplification of microwaves were therefore received with enthusiasm. These ideas lead to devices where a bunch of carriers is injected into a low-density drift region. The injected charge then travels with saturation velocities across this region. Whereas the injection of charge occurs somewhere during the positive peak of the r.f. voltage swing, charge drift continues for almost the whole period of the microwave oscillation. When the r.f. voltage swing turns positive again, the drifting carriers must reach the end of the drift region. During charge transit the device current remains correspondingly high, even for the negative part of the voltage swing. A high current for a negative r.f. voltage means a negative resistance, and microwave power is delivered to the microwave circuitry.

Carrier injection can take place (a) due to tunnelling, (b) at an injecting contact or junction (for a punch-through transistor, or for a metal-semiconductor-metal structure, the so-called baritt diode) and (c) by avalanching (the impatt diode). So far, option (b) has resulted in a range of impressive low-noise results, such as 10 dB for around 7·5 GHz amplification.[8] Impatt diodes are now commercially available for quite high c.w. powers, and operation at well above 100 GHz has been achieved. More than 4 W c.w. with 5–9% efficiency at X-band from a single diode mounted in a properly-designed cavity circuit has been reported. Particularly at frequency ranges above 50 GHz, the advantages of impatt diodes are greatest. Figure 5 indicates the state

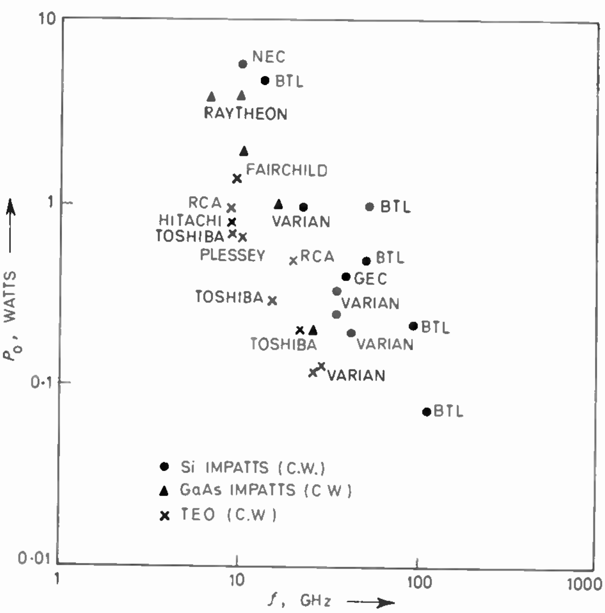


Fig. 5. Output power P_0 as a function of frequency f for a selection of the best results obtained so far with c.w. impatts and transferred-electron oscillators.

of the art of impatt oscillator power performance as a function of frequency. Theoretically, the maximum efficiency is 15% for Si impatts and 23% for Ge and GaAs impatts. As the avalanche breakdown is inherently noisy, relatively high noise powers have to be accepted. For materials with similar ionization rates for electrons and holes such as Ge and GaAs, the avalanche noise power is lower than in Si, which has different ionization rates for each carrier type.[9]

Numerous further developments have led to improved performance. A double-drift structure by ion implantation makes use of both types of carrier for the instability process.[10] The advantages are doubled output power and increased device impedance. The latter point is of value regarding complex-conjugate matching of these microwave devices to the circuitry, because impatts exhibit an unfortunately low impedance. Thus 180 mW at 92 GHz with 7·4% efficiency, and 1 W at 50 GHz with 14·2% efficiency were obtained. Impatt diodes with a Schottky junction have the advantages of better heat sinking (the metal employed has a higher thermal resistivity than that of semiconductors) and a smaller series loss resistance. Schottky-barrier GaAs impatts gave 4 W c.w. at X-band and 500 mW at 15–20 GHz with an efficiency as high as 8%.[12] A new, high-efficiency mode of avalanche diodes, the trapatt mode,[13] has been obtained in Si and Ge. A high-density of excess charge is created in the total depletion layer by an avalanching shock wave, which moves with a speed which is faster than the saturated drift velocity v_s This wave can be created if the driving current density is larger than $v_s e N_d$ (N_d is the ionized donor concentration). The terminal voltage across the diode then goes essentially to zero: this is the charging cycle. With the subsequent process, during the 'recovery cycle', the plasma is gradually removed by the recovering current I_R. The drift velocity of the trapped-charge plasma is smaller than v_s, as the rest field in the drift space is rather small. By choosing I_R, the recovering time can be fixed. In order to trigger trapatt oscillations, a very large impatt generated voltage swing is applied repeatedly to the diode. This is achieved by building a cavity which traps impatt frequencies, but which permits the passage of the trapatt wave (which is subharmonically related to the impatt mode) via a low-pass filter out of the cavity to a matched load. So far c.w. trapatt oscillations have only been obtained for frequencies around 4 GHz, although Evans[14] predicted c.w. output above 25 GHz with a reasonable efficiency of around 30%. The main feature of trapatt devices is the extremely high efficiencies predicted and achieved (even around 60%).

7 The Transferred Electron Effect

The trapatt mode of the last Section represents already an active microwave device, where a substantial part of the semiconductor volume rather than a narrow junction contributes to the instability. In the past there were many interesting proposals to employ various plasma phenomena in the bulk of semiconductors for microwave generation and amplification (for a review see chapter IV of ref. 15). The transferred-electron instability has been the most successful one so far.

The semiconducting crystal has to have such an energy-band structure that for low electric fields the charge carriers are in a low-effectiveness-mass state, whereas for fields above a threshold value the carriers are transferred by various scattering processes into an energy valley of the momentum space where its effective mass is increased. As the mobility is inversely proportional to the charge-carrier mass, a decrease in current occurs for increasing voltage, i.e. an effective negative differential resistance. This effect is very pronounced with GaAs and InP. GaAs devices have been commercially available already for several years now.

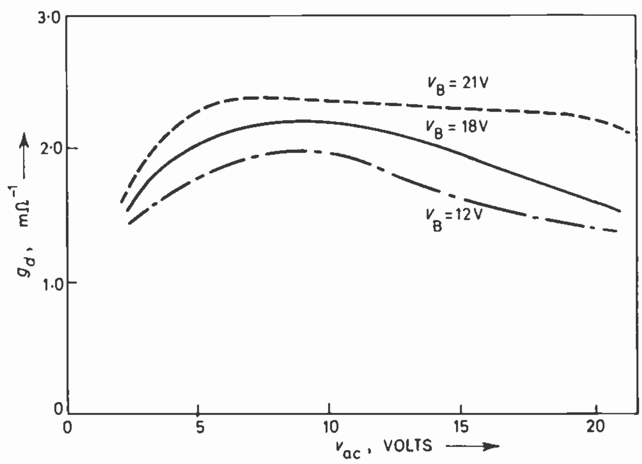


Fig. 6. Negative conductance g_d of Gunn diode vs. r.f. voltage amplitude v_{ac} for several bias voltages V_B (frequency of oscillation 9·7 GHz; low-field resistance R_0 5·5 Ω; device capacitance at $V_B = 14{\cdot}2$ V is 1·06 pF; cathode is n⁺ GaAs).

The space-charge-wave phenomena inside the semiconducting crystal with electron-transfer mechanisms in operation are complex with often large and rapid changes of the electric-field pattern between the two contact electrodes. These variations of electric field strongly affect the electron transfer process. As electrons are heavy at places of high field, travelling and stationary carrier-accumulation and depletion layers occur. The field pattern usually changes periodically at about the electron transit time frequency, and if a resonant cavity with a resonance frequency in the neighbourhood of this natural oscillation frequency is applied, the device delivers microwave power because the field changes appear as current changes at the device terminals. Computer simulations have shown that for the usual conditions of the oscillating devices, when the ionized donor density times interelectrode length product is about 10^{12} cm^{-2}, the various instability modes of the electron-transfer semiconductor plasma strongly depend on doping and mobility profiles. Modes can even change during switch-on transients or for increased bias voltages. This information can be obtained experimentally by determining the negative conductance as a function of microwave voltage amplitude across the diode, $g_d(v_{ac})$.[16] By applying a bias voltage in the form of a voltage step and measuring the transient wave, these data can be found relatively easily.

Numerous commercial diodes of different manufacturers have been measured in this way, and Figs. 6 and 7

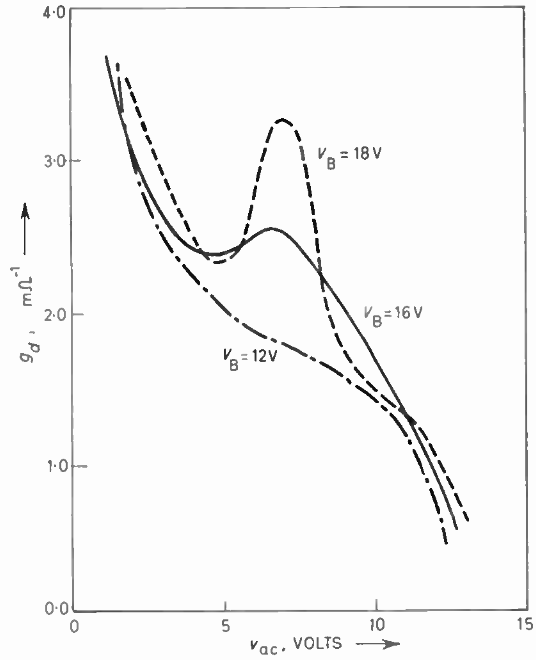


Fig. 7. g_d vs. v_{ac} for several V_B values (frequency 10·5 GHz; $R_0 = 9{\cdot}2\ \Omega$: device capacitance at $V_B = 14{\cdot}2$ V is 0·35 pF; cathode is a metal alloy).

show two typical results. Figure 6 gives a g_d value which remains high and almost constant over quite large v_{ac} ranges. Obviously high output powers and efficiencies are achieved. As g_d is small for small v_{ac}, the switch-on time is rather long. Small fluctuations in load conductance can cause large output amplitude changes.

Figure 7 shows a different type of diode where g_d is very large for small v_{ac} but falls off rapidly for increasing v_{ac}. This gives very short switching times but small efficiencies. Interestingly, for higher bias voltages V_B, a second peak sets in at around $v_{ac} \simeq 7$ V. Here a different space-charge mode appears. Such a second peak can be employed for memory applications, because a suitably selected load conductance gives then two stable operating points.

It is possible to correlate the experimental findings with computer-simulation results so that a clearer picture regarding the actual space charge mode present in a given device can be obtained. This can ultimately lead to guide-lines regarding the optimum choice of fabrication technology for devices of particular applications.

Figure 5 gives examples of the best performance, achieved at the present time with reliable transferred electron oscillators (t.e.o.s).

8 Conclusions

One had to advance a long way to make reliable microwave devices commercially available. This progress has been achieved over a period of only slightly more than ten years. Techniques had to be adopted which are similar to the solutions found earlier when microwave tubes were developed where transit times are incorpoated into the instability process. This meant that one had to abandon the simple transistor structure and therefore the devices described in the later part of this paper are very different from transistors. Mixing, parametric effects, tunnelling, carrier-injection transit-time devices and non-linear charge carrier transport such as the transferred-electron change of the effective electronic mass are successfully employed to generate and process microwave frequencies even up to 300 GHz. Because of the small component size and the prospect of long life times, entirely new applications become possible such as the overmoded-waveguide communication system covering 30–110 GHz, which is being experimentally tested now by several countries, or the cheap and small radar monitoring units for industrial and consumer markets. More applications have still to be found in order fully to exploit the opportunities these new devices offer.

There is, of course, a challenge now to conquer the submillimetre wave region with semiconductor devices. So far this area is covered by large-sized far-infra-red lasers involving gaseous-plasma chambers, except for the optical or near-infra-red semiconductor-junction laser. These frequencies will then also become more easily available for various engineering applications, where one has so far only the idea of communication via fibre waveguides.

9 References

1. Read, W. T., 'A proposed high-frequency negative resistance diode', *Bell Syst. Tech. J.*, **37**, pp. 401–46, 1958.
2. Johnson, E. O., 'Physical limitations on frequency and power parameters of transistors', *RCA Rev.*, **26**, p. 163, 1965.
3. Bächtold, W. 'Q-band GaAs f.e.t. amplifier and oscillator', *Electronics Letters*, **7**, p. 275, 1971.
4. Caulton, M., *et al.*, 'Generation of microwave power by parametric frequency multiplication in a single transistor', *RCA Rev.*, **26**, p. 286, 1965.
5. Irvin, J. C., *et al.*, 'A composite varactor for simultaneous high power and high efficiency harmonic generation', *IEEE Trans. on Electron Devices*, **ED13**, p. 466, 1966.
6. Kinoshita, Y., *et al.*, 'An 18 GHz double tuned parametric amplifier', *IEEE Trans. on Microwave Theory and Techniques*, **MTT-18**, p. 1114, 1970.
7. Archer, J. A., 'Improved microwave transistor structure', *Electronics Letters*, **8**, p. 499, 5th October 1972.
8. Bjorkman, G. and Snapp, C. P., 'Small-signal noise behaviour of companion p^+np^+ and $p^+n\nu p^+$ punch-through microwave diodes', *Electronics Letters*, **8**, p. 501, 5th October 1972.
9. Gummel, H. K., *et al.*, 'A small-signal theory of avalanche noise in impatt diodes', *IEEE Trans.*, **ED-14**, p. 569, 1967.
10. Scharfetter, D. L., *et al.*, 'Double drift region ($p^+pn\ n^+$) avalanche diode oscillator', *Proc. IEEE*, **58**, p. 1131, 1970.
11. Seidel, T. E., *et al.*, 'High-power mm wave impatt oscillator with both hole and electron drift space made by ion implantation', *Proc. IEEE*, **58**, p. 1135, 1970.
12. 'GaAs impatt extend c.w. power to 4 W', *Microwaves*, p. 48, February 1971.
13. Evans, W. J., 'Circuits for high-efficiency avalanche diode oscillators', *IEEE Trans.*, **MTT-17**, p. 1060, 1969.
14. Evans, W. J., 'Computer experiments on trapatt diodes', *IEEE Trans.*, **MTT-18**, p. 862, 1970.
15. Hartnagel, H., 'Semiconductor Plasma Instabilities' (Heinemann, London, 1969).
16. Kawashima, M. and Hartnagel, H., 'New measurement method of Gunn diode impedance', *Electronics Letters*, **8**, p. 305, 15th June 1972.
17. 'Power transistors climb to 4 GHz with 5 W output', *Microwaves*, p. 60, March 1972.

Manuscript received by the Institution on 25th October 1972.
(Paper No. 1498/CC162)

UDC 621.382: 621.39

The expanding role of semiconductor devices in telecommunications

J. R. TILLMAN,
D.Sc., F.Inst.P., C.Eng., F.I.E.E.*

SUMMARY

The junction transistor made its first impact on telecommunications in transmission systems, then entirely analogue, effectively by replacing thermionic valves. Later developments of that transistor have made possible systems enabling over 10 000 telephone channels to be carried on one 9·5 mm diameter coaxial tube. Submarine cable systems have also benefited from the development of highly reliable wideband transistors, with a 1800-circuit single-cable system now in production for intercontinental working. Recently, telecommunications satellites, owing their success electronically to several different classes of semiconductor device, have challenged repeatered cables, offering even higher capacities.

Exploitation of the transistor as a logic or switching device followed the early analogue applications, with 24-channel digital transmission systems using pulse code modulation. The control of switching systems was soon demonstrated, with intensive follow-up, even though metallic contacts through which the audio signals are routed in public telephone exchanges have not yet been supplanted.

Some systems under development demand special devices, e.g. solid-state sources for frequencies up to 100GHz for both guided and new free-space microwave systems. Guided optical communications make desirable a long life, room-temperature-operated, high duty ratio gallium arsenide laser. Integrated circuits will find a very wide range of use throughout.

**Post Office Telecommunications Headquarters, Research Department, Dollis Hill, London NW2 7DT.*

1 Introduction

The transistor was born and nurtured in the Bell Telephone Laboratories (BTL), and its development led, there and elsewhere, to that of many other semiconductor devices. It would therefore have been ironic had telecommunications not benefited as much from these developments as did, perhaps more obviously, some other important branches of electrical engineering, such as digital computing and consumer electronics. But the impact on telecommunications, widespread and radical as it now is and even more so to be in the future, was slow to build up during the 1950s. Very wisely the BTL concentrated almost entirely on improving the devices in the first few years after 1948, rather than on pursuing all possible applications of the earliest devices whose performance we now recognize as having been very limited. But they certainly showed their faith in the new device by using point-contact transistors to amplify signals from point-contact photodiodes in a card translator,[1] designed to take the telephone address of a call and to determine how to advance the call towards that address; translators were installed at some control switching points.

Elsewhere, telecommunications laboratories refrained from any serious attempts at applications of transistors until junction devices became freely available. These laboratories were very busy developing new systems, or much improved versions of old systems, made possible in part by war-time advances in electronics and the accompanying improvements in both general receiving and microwave thermionic vacuum devices. Telecommunications networks needed rapid expansion particularly to meet the much increased demand for national and international trunk circuits. The multichannel systems chosen for development demanded, along with other component parts, improved modulators, carrier supplies, line or intermediate-frequency amplifiers and channel amplifiers, and of these only the modulators ever depended on a solid-state device—the copper oxide diode. Success attended these developments and wideband coaxial systems and microwave radio relay systems, carrying analogue telephony and television signals, came to dominate long-distance transmission in the industrial countries; the microwave systems used either triodes with very closely spaced electrodes or travelling wave tubes (t.w.t.s) to produce the necessary radio frequency power.

Systems using coaxial cables and submerged repeaters, lying on the seabed, were just being introduced to supplement radio links often very limited in capacity and of variable quality, between countries separated by sea distances of more than about 50 km. Rapid progress to systems using many repeaters in series was marked particularly by the first transatlantic system (TAT1) brought into service in 1956; the main crossing from Scotland to Newfoundland used two cables, unidirectional amplifiers and specially developed thermionic valves to give 36 circuits. In 1961 a more revolutionary system, using one cable only and valves of higher performance, between Scotland and Canada afforded 60 4 kHz circuits, raised later to 80 circuits with the adoption of 3 kHz channelling for long submarine systems. Similar 80 and 160-circuit systems were later used in the Atlantic and Pacific to form a considerable network. Even so traffic threatened to demand many more circuits, and an attempt was made to meet the threat with the aid of another stage of improvement to the valve. Although successful in making possible a 360-circuit system, only one long distance route was laid before the transistor took over the role of the valve in submarine telephony.

As the capacity of the main trunk networks increased, the need for greater capacity on shorter routes became very pressing. One quick way of meeting this need was adopted in the USA in the early 1950s. It called neither for additional cables, nor the limited amount of carrier working which the existing audio cables might have allowed, at considerable cost for terminal equipment. These many audio cables had almost always been used four-wire. They were now converted to two-wire working, with transmission loss made good by the insertion of a negative impedance amplifier, usually near the middle of each two-wire circuit. A cable's capacity could thus be doubled, crosstalk permitting.

New signalling systems, using signals which could be transmitted wholly within the speech band were also necessary, to facilitate operator—and later customer—dialling over long distances. They too were forthcoming, based essentially on low-power receiving valves.

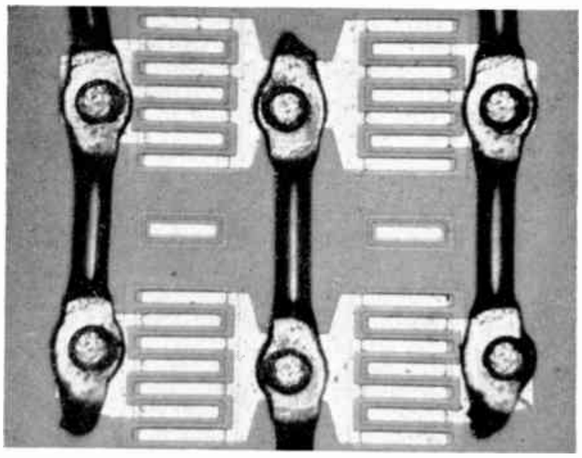
When the transistor was in its infancy, the major practices in telephone switching, that very big field of telecommunications, were not turning towards electronics, except in a few laboratories where ideas for major involvement of electronics were still tentative and in need of much experimentation before well-considered proposals could be made. Electromechanical components, such as relays, two-motion selectors and crossbar switches, held sway. Their overall direct replacement by solid-state devices did not seem possible, but their control seemed within the reach of existing electronics. Alternatively entirely new systems might be conceived. The early work on digital computers was leading to some switching or logic techniques of possible value, and there were the incentives of speeding-up many of the functions of telephone exchanges and of eliminating metal-to-metal contacts whose wear, erosion and corrosion had long presented serious maintenance problems. While it was not possible to predict whether evolution or revolution in exchange design would become the main objective, progress was bound to be affected, constrained indeed, by the strong—if not indeed economically essential—need to maintain full interworking between old and new systems with as little interfacing as possible.

2 The First Phase of Penetration, in an Analogue Context

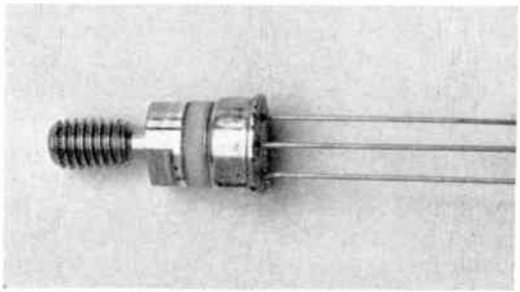
The scene described above was not seriously affected until the junction transistor became available, at steadily reducing prices, with increasing bandwidth and some assurance of reliability. It then quickly became obvious that there was just as ready a takeover situation in telecommunications for the transistor to replace the valve, as elsewhere; and this replacement took place largely without any fundamental systems changes having to be made. Dispensing with heater supplies was an economic benefit, but smallness of size initially counted for little.

The takeover began at the audio end of the frequency spectrum. No serious difficulties arose in designing repeaters, two-wire (negative impedance)[2] or four-wire, and channel amplifiers with adequate gain stability, linearity and low noise figure. Wide-band amplifiers for carrier systems had to wait a few years for transistors with much greater gain-bandwidth products (i.e. $f_T >$ 100 MHz). The wide spread of gain factor (grounded-emitter short-circuit current gain h_{fe}) proved no obstacle, and fears that some excess phase shift might prevent the application of sufficient negative feedback to ensure adequate suppression of intermodulation products, without reducing unduly the margins against loop instability, proved groundless. Today therefore we have, in wide use, systems carrying 5, 16 and 45 supergroups on coaxial tubes of diameters ranging around 4·4 and 9·5 mm.[3] In many systems the line amplifiers are mostly buried, with power feeding from, say, every eighth repeater location, which is above ground. A 60 MHz cable system is currently under development in several countries. In the UK it will use air-spaced-coaxial tubes with a diameter of 9·5 mm. Eighteen such tubes in one sheath—nine for each direction of transmission—give the whole system a capacity of nearly 90 000 4 kHz circuits, plus some spare capacity. The line amplifiers will use transistors with gain-bandwidth products (f_T) of about 2 GHz, and will fully meet the stringent noise and intermodulation requirements set for main trunk routes.

Submarine telephony fell to the transistor in the mid-60s. The UK was early in the field[4] with several routes up to 1500 km in length, using a silicon planar transistor of much improved reliability, to permit 8 supergroups to be carried both-way over one cable. The ATT, using a germanium mesa transistor developed by BTL, later laid TAT5 between the USA and Spain to carry over 800 3 kHz circuits.[5] Meanwhile developments by the UK Post Office[6] of a wider band, silicon transistor of very high reliability made possible a capacity of 1800 3 kHz circuits, such as for the system designed by STC to be laid in 1973/4 between the UK and Canada (Fig. 1). The quantitative assessment of the reliability of these transistors is an extensive exercise.[7] It can be achieved by scientifically based, accelerated ageing tests consisting of very severe overstressing of a randomly selected majority of an already screened production batch. At intervals during the overstressing, which is usually achieved by way of much increased ambient temperature with normal biasing, the electrical properties are repeatedly re-measured and the results carefully analysed for changes whether systematic, continuous, random or otherwise. These 'assessment' transistors may well be aged for the equivalent of 100 years and more of normal working life. They must indicate the necessary survival rate (e.g. not more than 1 failure in 2000 over a period of 25 years), extrapolated to normal conditions of use and with failure criteria which are very stringent. Only then will the remaining, minority, fraction of the batch, not over-stressed in this way, be considered for service use.



(a) Showing the surface geometry of an output transistor, capable of dissipating 1·5 W consistent with extremely high reliability and making possible submarine cable systems carrying 1800 circuits. The metallized finger width is 12 μm and the eyelet-bonded wires (the two outers for the emitter and the other for the base) are 38 μm in diameter.



(b) Illustrating the encapsulation of the transistor, including the electrically isolated, brazed, stud mounting of low thermal resistance. The diameter of the header from which the lead wires emerge is 9 mm.

Fig. 1. A silicon planar transistor for submarine telephony.

But while submarine cable systems of even greater capacity are under development in several countries, based on new designs of transistor of increased bandwidth, competition has been growing, in the field of intercontinental telecommunications, from earth satellites. By 1960 the suitability and reliability of components, particularly semiconductor devices, were under scrutiny for use in low-orbit satellites.[8] Early experience was promising, and further development, not forgetting one particularly interesting case history,[9] elucidated and solved residual problems. In the main the usages of semiconductor devices are conventional, both in telecommunications satellites (now predominantly in circular orbits, i.e. geostationary), and in the Earth stations. But three deserve mention. The first is that of tunnel diodes, which had not previously played an important role in telecommunications systems. The transponder of the *Intelsat IV* satellite,[10] the first of which was launched in January 1971, receives signals in the frequency band 5·9–6·4 GHz at very low level which must be amplified with minimum loss of signal/noise ratio, and without the availability of any coolant to ease the problem. Tunnel diodes (in germanium), despite the difficulties they bring by being two-terminal devices, are currently chosen in preference to any transistor device; further developments may however change this choice—both for systems using the same frequency band and new systems using higher frequencies. The front-end low-noise amplifier (giving a gain of about 12 dB) is followed by a mixer, with a semiconductor local oscillator, converting to the 3·7–4·2 GHz band; a second tunnel diode amplifier, and a low-level travelling wave tube follow before the signals are separated into r.f. channels and each amplified by high power t.w.t.s. to be transmitted back to earth (Fig. 2). Normally *Intelsat IV* will carry 7500 telephone channels, with additional capacity for television. The second use—already well-established for space vehicles—is that of silicon solar cells to power the satellite. The third is the use in the receiver of the Earth station of a cooled parametric amplifier, (using a gallium arsenide varactor diode) capable of handling the band 3·7–4·2 GHz at a received power which may be as low as 10^{-13} W. By cooling to about 18 K, noise temperatures of 15–20 K can be obtained with gains of 30 dB, before the signals are further amplified, distributed by branching networks to down converters, amplified at some intermediate frequency and demodulated to base band.

3 The Growth of Digital and Switching Applications

The first major applications of transistors in telecommunications made use of their good performance as analogue devices over a wide range of frequencies. But simultaneously devices of basically the same design were the key to the second generation of digital computers. Here however it was the excellent suitability of the transistor as a logic and switching element which was being called into play, deriving from the clearly defined, low power-consuming end states, and the high speed with which switching between the two states could be effected, without any serious regard to the linearity of that transition. In association with other components, mostly diodes and resistors, they formed the basis of many forms of logic gate, bistable circuit, counter, adder, etc.

Twenty years ago, telecommunications had few ready-made applications for these, or similar, types of logic and

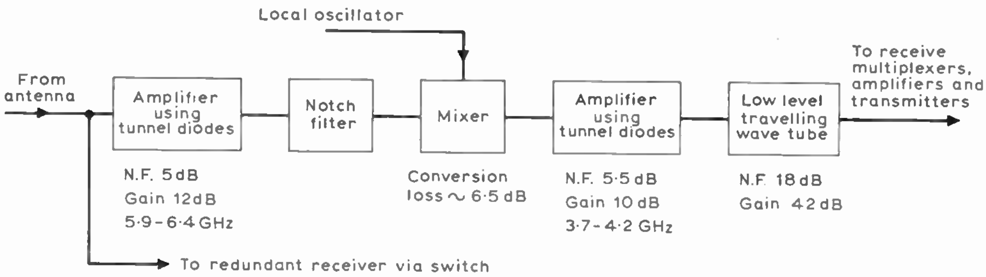


Fig. 2. Transponder of *Intelsat IV*: receiver block diagram.

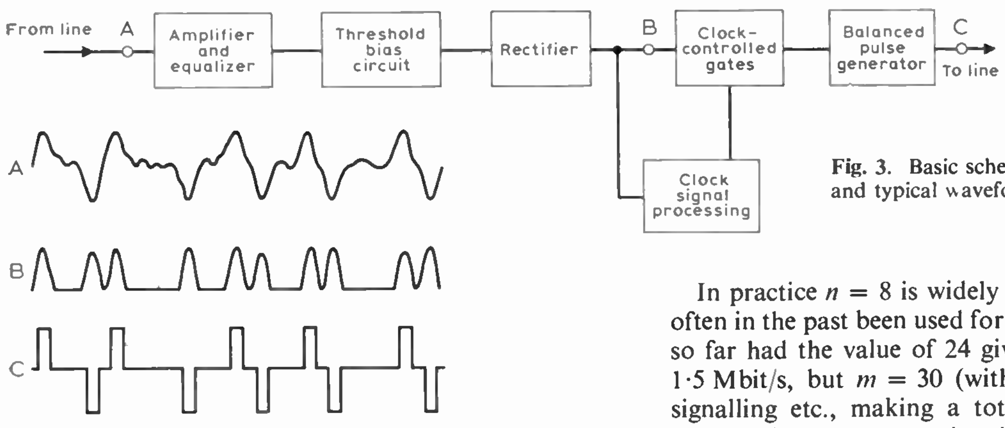


Fig. 3. Basic schematic of a p.c.m. regenerator, and typical waveforms at key points.

switching circuits. But their use in computers prompted the study of new systems which could exploit them. One such system revived an idea, first proposed in essence in 1938,[11] for using pulse code modulation along with time-division multiplexing to increase the circuit carrying capacity of existing multipair cables normally carrying only audio circuits.

In each direction of such a system, the separate audio signals of the m circuits to be multiplexed are first filtered to reject components above 4 kHz. They are then sampled at a frequency of 8 kHz (i.e. once every 125 μs). The samples (in the shape of amplitude modulated pulses) are time-division-multiplexed and passed to an analogue-to-digital converter. Here their amplitudes in turn are measured against a set of reference levels (equivalent to 2^n amplitudes) compressed rather than uniform, and each is quantized to the nearest of such levels, which can be expressed as a n-digit binary number. The n digits representing each sample are transmitted to line by the presence or absence of each of n successive pulses, contained within a time slot—repeated every 125 μs for each of the m inputs—of duration $(125/m)$ μs. The resultant complete pulse train can be regenerated at regular intervals along the route (spacings of about 2 km are common) with negligible error rate even though the signal/noise ratio drops well below (but clearly not too grossly so) that tolerable for conventional amplitude-modulated systems (Fig. 3). The audio signals can be recovered very faithfully at the receiving terminal by processes reciprocal to those at the transmitting end.

In practice $n = 8$ is widely used, though one digit has often in the past been used for signalling purposes. m has so far had the value of 24 giving a bit rate of just over 1·5 Mbit/s, but $m = 30$ (with two extra time slots for signalling etc., making a total of 32) is being recommended in some countries, including Western Europe, giving a bit rate in excess of 2 Mbit/s. Field trials of a system were held in the USA in 1960 and large-scale installation began a few years later.[12] The UK followed with a similar system a few years later, as did Japan. First equipments, both terminal and intermediate, used many discrete transistors, but later installations made very considerable use of standard bipolar integrated circuits. Low circuit costs and a very good quality of transmission have been achieved.

The success with p.c.m. systems with $m = 24$ (or 32) on multi-pair cables over short routes has led to much consideration of a hierarchy of systems (to be compared with 12, 60, 240 etc., channels of f.d.m. systems). A planned hierarchy enables easy assembly on, and distribution from, high capacity routes. Transmission media for systems with the increasingly higher values of m will initially be largely coaxial cables. These systems will be suitable not only for telephony, but for the transmission of person-to-person television services, a channel for which might require 8 Mbit/s, or of 625-line colour television requiring nearly 120 Mbit/s, though both of these rates may well be reduced by future studies of bit rate conservation. Considerable development has gone into many aspects of these high capacity systems, which seem certain to become serious competitors to the more conventional wideband analogue systems. The semiconductor elements required are already available either in discrete or integrated form.

An interesting example of a highly specialized switching equipment, made possible by the transistor and developed

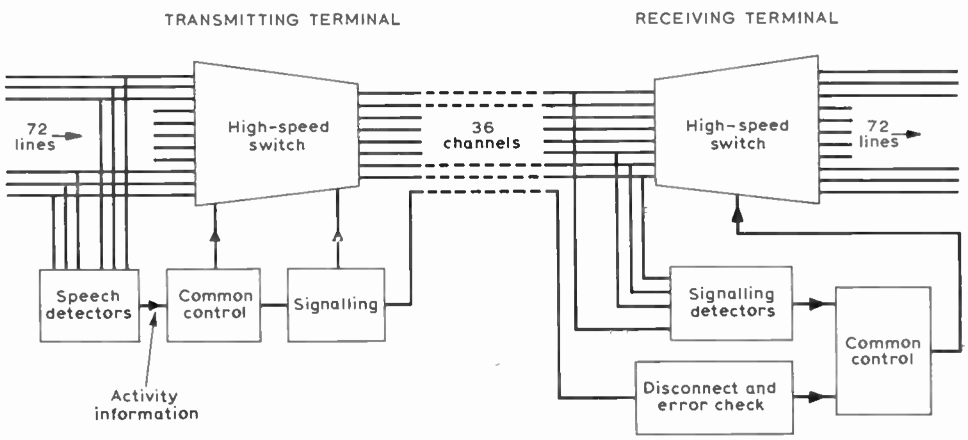


Fig. 4. Basic schematic of TASI system. The numbers of lines and channels, which are frequency division multiplexed, refer to the system designed for TAT1. The Figure shows only half a system, namely that handling the West-East direction of transmission. The other direction uses similar equipment, East transmitting and West receiving.

at much the same time as the first p.c.m. system, is that of the TASI (Time Assignment Speech Interpolation) system (Fig. 4).[13] This system seeks to use the considerable idle time in four-wire telephone connexions—nearly two-thirds in each direction in most calls—for the interpolation of additional speakers. Studies showed favourable economics over long routes only, and the first usage was on TAT1 (mentioned earlier), only a few years after its laying. The TASI equipment consists essentially of a transmitter and receiver at each end of the long distance route (e.g. at London and New York). They are inserted between the normal terminating plant and the local, national or transit lines to be served. The signals are treated as four-wire. Each transmitter is complex and it includes detectors, responding to every burst (talk spurt) of speech incoming from the lines, and a high-speed switch which converts incoming audio signals to amplitude modulated, time-division-multiplex pulse trains. The switch steers pulse trains to idle channels where they are reconverted to audio. These channels are selected by a common control equipment which is continuously supplied with information about the activity of the lines to be served in order to decide the optimum use of the channels. Signalling, also determined by this control, is by multi-frequency tones sent immediately preceding each talk spurt to enable correct operation at the receiving end of the long route. Queues may occasionally form, competing for idle channels as they become free. But the system could be designed so that, for a doubling of the number of conversations that could be carried, clipping of the start of talk spurts is barely noticeable and freeze out (loss of talk spurt) very rare indeed.

Each terminal of the original design contains several thousand transistors and tens of thousands of other components—principally diodes and resistors—whose mounting occupies 2500 printed circuit boards. Their combined functions today could, almost entirely, be performed by integrated circuits occupying much less space. But the use of TASI seems unlikely to be extended to the newer much higher capacity submarine systems, or to national networks where growth is preferred to be by additional cables or radio relay systems.

These two examples of systems made possible by the excellent logic/switching properties of transistors apply to the need for increased capacity of transmission systems—one for short routes, the other for long routes. But a third exploitation of these properties, to improve switching systems, seems likely to prove a very large outlet for the use of semiconductor devices.

Electronic switching systems, or electronic telephone exchanges, dispensing in part or whole with electromechanical devices, attracted increasing study after 1950. Attention was given both to the prospects of supplanting the metallic contacts which have to be made or broken in setting up or clearing down the conducting paths to carry the conversations, and to the control of the functions of the exchange—in accordance with the signals received by it indicating connexions to be made, the facilities to be given, the available paths through the exchange, the outlets available, etc. Some early attempts were made to use p-n-p-n devices, diodes or triodes, as crosspoint switches, but neither type gained any serious foothold, for reasons including inadequate conductance when 'closed' and inadequate isolation between the controlling and controlled circuits. The limitations may not have been absolute, but to overcome them imposed intolerable economic penalties. The control functions seemed more easily achievable technically, but the capital cost was high so that the economic case rested primarily on reduced maintenance, supplemented by the additional facilities offered and services made possible.

Many systems ideas were studied in detail before any design went into production. One major field trial in the USA,[14] designed by BTL, used transistors extensively in very simple gates, flip-flops etc. as control elements, with gas discharge devices as crosspoints. But the first system to be produced for service use was the No. 1 ESS,[15] designed by the same laboratory and described as the largest single development project ever undertaken by that laboratory for the Bell system. It reverted to the use of metallic contacts for crosspoints, in the shape of sealed reed relays. Their operation is controlled electronically along with many other functions of the exchange by a central processor (duplicated to prevent catastrophic breakdown should a fault occur), instead of by the multiplicity of control units of conventional electromechanical exchanges.

Stored program control (SPC) is incorporated thereby easing manufacture and testing in the factory, and improving flexibility in use, e.g. allowing for sizes between 8000 and 65 000 lines; program storage is in semi-permanent memories along with translation information peculiar to the particular exchange. A temporary memory stores information about calls in progress—obtained from scanners continuously at work probing the lines and connexions. The central processor decides the course of action to be taken by much of the other hardware and acts through signal distributors. SPC puts a big responsibility on adequate software, but dispenses with some hardware and reduces physical modifications as circumstances change. Computing techniques abound, but one philosophy of the exchange is diametrically opposite to that of a digital computer. For the exchange, a small (even undetected) mistake in operation is tolerable; but a shut-down however brief is intolerable. For the computer a small (undetected) mistake may well be intolerable, but a shut-down—though irksome—is not intolerable. (This situation is less tolerable with multi-access working.)

The No. 1 ESS uses many semiconductor devices, but only a few types (see Table 1 which ignores three other types used very sparingly only). Since plans were laid to connect about 10^6 lines annually to this type of exchange, the number of devices so needed, to accompany about 15×10^6 ferreed crosspoint switches (each containing two reed relays), is very considerable.

In the UK, after some attempts to pursue all-electronic designs, which met with difficulties, the reed relay was again adopted for the crosspoints. The first exchanges to go into service, the TXE2,[16] accept only 2000 lines each—later extendable to 4000 lines. Control is by way of registers, but semiconductors do not dominate all their functions, some reed relays serving here as well. Over 200 such exchanges are now installed. Production is starting of a larger exchange with reed relays again serving as crosspoints, but leaving almost all the control functions to junction devices.

Table 1
Semiconductor devices for No. 1 ESS

Type	Function	Approximate number required for an installation serving 10 000 lines
Transistors	Logic	70 000
	Driver	2000
Diodes	Reference	4000
	High current	5000
	Medium current	25 000
	Logic	125 000
	Level shifter	45 000
	Voltage regulating	1000

Developments in large-scale integration could well assist a trend away from total concentration of control functions in large central processors, which have to be fully—and hence expensively—protected against faults which could otherwise result in complete system failure. More distributed forms of control may be preferred, wherein peripheral faults have only a limited effect on service. The trend is facilitated by the concept of 'programmable logic', that is by the use of a general-purpose logic array which can be adapted to perform a specific function by a program stored in a microelectronic read-only memory. The integrated circuits for the general-purpose logic units can be standardized to take advantage of large scale production and so to reduce costs. The approach resembles the well-known technique of micro-programming, but with the microprogram removed from the central computer to the periphery.

These electronic systems are designed to switch voice signals in the same analogue form as did their electro-mechanical predecessors. But with short-haul p.c.m. links being increasingly used and with high-capacity trunk digital systems being developed to carry conversations in a digital mode, a new prospect arises. If much of an overall connexion is to consist of such links separated only by switching centres, would not switching in a time-division mode be preferable (more economic as well as giving better transmission) to demodulation of the p.c.m. signals to space division audio signals and reconversion to digital form after switching? Time-slot changes would have to be faced along with problems of synchronization (Fig. 5). The UK Post Office has shown in an experimental tandem exchange,[17] being given an extended field trial, that there are acceptable technical solutions to all the problems. Here, freed from such duties as passing ringing currents, the equivalent of the usual crosspoints can be achieved by semiconductor devices; indeed they are essential to achieve the operating speeds required.

The design of the exchange could in fact be looked upon as an exercise in the use of standard d.t.l. and t.t.l. integrated circuits (about 10 000 in all), some discrete transistors and diodes and several memories. Its success in the field augurs well for what will have to be more ambitious systems when the networks are carrying digital signals from sources other than telephony, of higher bit rates and different formats.

4 Further Penetration

The examples given in the two previous sections, of the coming of extensive use of junction devices, have been in the fields of transmission and switching. Other fields benefited as well, however, not only through improvements to long-standing systems or specific equipments, but in the implementation of new ideas. Because nothing exceptional is required of the devices utilized, except long life, only a very few examples in these fields will be mentioned here.

Signalling systems account for an important fraction of all investment in telecommunications, particularly with the expansion of long distance dialling facilities. The receivers of the in-band (voice frequency) systems have used discrete semiconductors for many years now. A recent design in the UK, intended for widespread use, will employ integrated circuits. The signalling tone (2280 Hz) is recognized, not by an LC circuit as has hitherto generally been used, but by an active filter incorporating a gyrator using two monolithic operational amplifiers in order to dispense with the need for a bulky inductor; subsequent processing of the signals after recognition, formerly performed by relay sets, will be carried out largely by low-power logic integrated circuits.

Many attempts have been made to simplify dialling—for the customer. The Bell Systems 'Touch Tone' calling was initially designed to be compatible with some electro-mechanical exchanges and used discrete transistors. The telephone incorporates a numbered (0–9) set of keys which when depressed cause the transmission of pairs of tones (one out of four and one out of three for instance) to identify each digit for reception at the exchange. A later

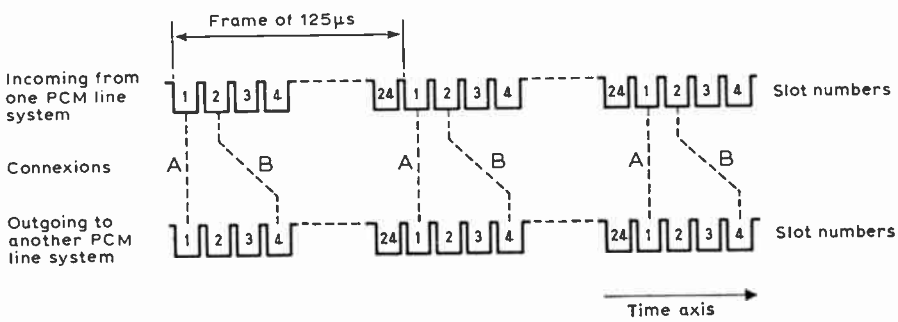


Fig. 5. Basic principle of time switching. Each slot, of duration 5·2 μs, can contain 8 bits (7 to define the sampled amplitude of the signal being conveyed and 1 for the signalling code).

version[18] designed for use with No. 1 ESS uses hybrid integrated circuits consisting of film circuits based on tantalum and its compounds together with beam-leaded active elements.

A demand for the transmission of data signals over a network in a flexible way, i.e. not tied to dedicated lines, began to grow ten years or so ago, with computer terminals and many more teleprinters coming into use. The development of interfacing equipment (data modems) to convert these signals into a form which can be carried on a voice channel, has offered to customers—first in the USA, later in the UK and more recently in some other countries—a choice of bit rate, mostly in the range 200–2400 bit/s. Higher rates are less in evidence, but recently, in the UK, a 48 kbit/s service has been offered, aimed at computer users, using the bandwidth of a conventional 12-channel (frequency division multiplex) group (i.e. 48 kHz). Clearly these low speeds make no serious demands on the devices available, compared with the faster digital systems mentioned in Section 3 and to be touched upon in the next section.

Private automatic exchanges are now obtainable with a large element of integrated circuits and, in at least one design, with four-layer semiconductor crosspoints. Repertory callers are coming into service which enable a customer to store several telephone addresses in his handset—anything from one address upwards, within reason. Each address can consist of up to the maximum number of digits ever likely to be used. Early models were very straightforward, e.g. with each address having its own dedicated key to be depressed, and simple transistor circuits to control the pulse train transmitted; later versions used tapes or cards to facilitate entry and modifications to the addresses stored. Current developments however are exploiting l.s.i. m.o.s. and semiconductor memories.

Telecommunications makes extensive demands on electronic instruments, many designed specially for equipment developments, for commissioning and for maintenance. Here again many benefits have resulted from the use of semiconductor devices, utilizing improved performance, greater reliability, low power consumption and small size.

5 Some Future Developments

Extensive as the use of semiconductor devices has now become in telecommunications, the future holds every promise of many new applications and further general penetration. The scope for integrated circuits—from standard ranges to custom-designed l.s.i. types—covers a wide range of speeds of signal processing. At the low end of the speed range the micro-power dissipation of such technologies as complementary symmetry m.o.s. (c.o.s.m.o.s. or c.m.o.s.) may prove a key factor, whereas at the high end the most advanced designs of today may prove inadequate.

Analogue applications will continue, within a wide range of bandwidth. A replacement for that long standing component of telephony, the carbon granule microphone, has repeatedly been sought to give improved linearity and better long-term stability. The best contenders for this role all lack efficiency however, making amplification essential before signals are transmitted to line. A low-power linear integrated circuit has a role here.

In the past, with notable exceptions in the Bell System, most of the devices used in telecommunications have become available as the result of prior work elsewhere, e.g. for computers or military electronic systems. The future could well see more developments specifically for telecommunications, as some systems now under study suggest. Good examples can be drawn from proposed millimetric and optical systems.

Long-distance transmission by waveguide[19] using a low-loss mode at millimetric wave frequencies demands an accurately circular guide laid very straight. A guide of 50 mm diameter will enable use of the band from 35 GHz to 100 GHz, with losses of no more than 3 dB/km in straight runs. For the digital systems favoured, individual channels within this band would have widths of between 0·5 and 2·0 GHz, capable of carrying 250–1000 Mbit/s, or more with multi-level modulation. Telephone administrations had discounted the practicability of such a system, despite its high potential capacity, in part because of a lack of suitable and reliable devices for the generation of the necessary carrier frequencies. But success in producing semiconductor devices for use at frequencies up to about 10 GHz gave a spur, both to establishing that suitable guide could be made and laid, and to extending the device work, mostly centred on impatt diodes, upwards in frequency. Some success has already been achieved all the way to 100 GHz, though adequate reliability has yet to be established. The regenerators required need the very fastest i.c.s available today. The Bell System has shown its confidence in this type of system by announcing its intention to have a 100 mile link in public service before 1980.

Terrestrial microwave radio relay links can use parts of the same frequency range and if they do so will need similar devices. Studies are already being made of how best to employ the frequency bands allocated to telecommunications, for city to city links and, at the higher frequencies, for local distribution. Hybrid microwave integrated circuits could find considerable use ultimately both here and in guided systems.

Another transmission system of potentially high capacity, now under study, uses optical signals guided along one of several possible forms of glass fibre of very low transmission loss (e.g. 10 dB/km) at the optical frequencies chosen.[20] One system of potentially high capacity requires monomode transmission along a fibre with a core of very small diameter ($\sim 2\ \mu m$) clad with material of slightly lower refractive index. The source can most conveniently be a GaAs laser, directly modulated at rates up to several hundred Mbit/s. Much effort has gone into making this type of laser suitable for use at room temperature, with a high duty ratio. The introduction of thin layers of gallium aluminium arsenide into the structure has reduced the threshold current, and hence power dissipation, by a large factor in the last two years; but life under the wanted operating conditions is still very low, demanding studies of the failure mechanisms with a view to their elimination. Silicon avalanche photodetectors will probably suffice as the low-level receiver, and regeneration of the optical pulse trains should present no new demands for semiconductor devices.

It was early recognized that person-to-person television would require a camera tube with a more robust target than the antimony sulphide type used in vidicons, and to this end the Bell Telephone Laboratories developed an alternative consisting of an integrated array of about 525,000 photodiodes within an area of 1·2 cm^2 on a slice of silicon.[21] The need for electron beam accessing remained with, in turn, the need for a vacuum enclosure and scanning facilities. Alternative arrays of photodiodes with scanning and read-out facilities more completely integrated on the silicon slice, such as are already being developed, would offer big advantages for this particular application, and possibly to other television services.

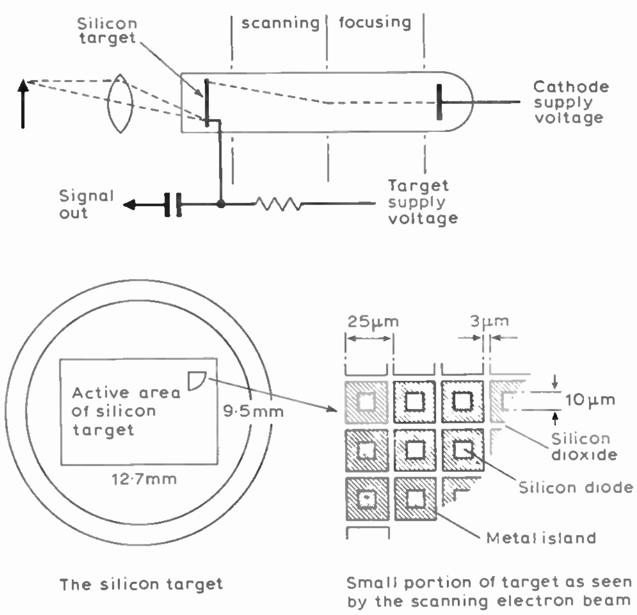


Fig. 6. A television camera tube with silicon-diode-array target. The dimensions given (bottom right) correspond to an array of 500 × 375 diodes.

6 Conclusions

The extensive penetration of semiconductor devices into telecommunications has certainly had, and will continue to have, a major impact on the designers and suppliers of equipments and systems, and on the public and private utilities re-equipping and expanding their total systems. But most customers of the services offered can have felt no such obvious impact. To them telephony, which accounts for so much of the services, has changed little in the last two decades. The resultant reduction in size of switching and transmission equipment now made possible, the lower power consumption, lower fault rate and generally greater signal handling capacity is largely hidden from them. The cost (and hence rental) of their terminals, and of making local calls, has been little affected and inflation has largely obscured the fall in rates for long distance calls.

The future could well change this scene. It will probably do so only slowly, other than for business use, if the customer spends no greater proportion of his income on telecommunications in the future than he has in the past. But it could change rapidly if he decided to spend as increasing a proportion as he has recently on television, i.e. since colour services were introduced. New facilities will be offered, almost always dependent on semi-semiconductor devices—particularly integrated circuits. Visual displays and data terminals could find uses in homes as well as offices; some of them need make no extraordinary demands on the networks. Person-to-person television does make serious demands—on switching in some countries as well as on transmission—and is therefore unlikely to be adopted widely and rapidly as yet. Information stores may well be provided, e.g. to allow a customer to key in his intended movements so that telephone calls made to his normal number will be redirected without either delay or realization by his callers. Accessing this store, or setting-up calls, by voice commands seems a long way off as yet, even with much discipline in the process, but when it does become possible integrated circuits will have a major role to play. Accurate prediction of the many new services possible and the dates for their introduction involve factors other than the purely technical however, and will not be pursued further here.

The telecommunications industries, manufacturing and operating, will continue to look for improved and new devices from the semiconductor industry. Few new devices which become established—other than those passing very high currents—will escape their attentions and most will find some ultimate role in their fields if the past is any guide. At the moment no striking new physical effects or new materials seem likely to satisfy some outstanding or easily foreseen needs, e.g. for solid-state crosspoints of fully adequate performance, cheap large-scale-integrated electrically-alterable non-volatile memories, low-voltage luminescent panels and simple fast operating optical switches and deflectors. But as the search proceeds for solutions to these device requirements and to problems of better signal transducing, processing and transmission, there may well be uncovered radical new effects such as led to field effect devices, tunnel diodes, transferred electron devices and the semiconductor injection laser. It may indeed lead to something as outstandingly novel as the junction transistor which in giving birth to a new industry is exercising a profound influence on telecommunications.

7 Acknowledgments

The author wishes to thank many colleagues, too numerous to mention by name, for much factual information.

Acknowledgment is made also to the Director of Research of the Post Office for permission to make use of the information contained in the paper.

8 References

1. Hampton, L. N. and Newsom, J. B., 'The card translator for nationwide dialling', *Bell Syst. Tech. J.*, **32**, p. 1037, 1953.

2. Bonner, A. L., Garrison, J. L. and Kopp, W. J., 'The E6 negative impedance repeater', *Bell Syst. Tech. J.*, **39**, p. 1445, 1960.

3. See, for instance: Prior, H. T., Chapman, D. J. R. and Whitehead, A. A. M., 'The application of transistors to line communication equipment', *Proc. IEE*, **106**, part B, p. 279, 1959. Norman, P. and Howard, P. J., 'Coaxial cable system for 2700 circuits', *Electrical Communication*, **42**, p. 499, 1967.

4. Myerson, R. J. W., 'St Margaret's Bay—La Panne 420-circuit submarine-cable system with transistor-type submerged repeaters', *Post Office Elect. Engrs. J.*, **60**, p. 1, 1967–8. (Because this system required only two repeaters it proved possible to find transistors already available with acceptable reliability; all later systems have needed specially developed devices such as those described in Ref. 7.)

5. 'SF submarine cable system', *Bell Syst. Tech. J.*, **49**, pp. 601–798, 1970. (Ten papers in all.)

6. 'A 12 Mc/s submerged-repeater system', *Post Office Elect. Engrs J.*, **59**, p. 199, 1966–7.

7. Corke, R. L., Hare, A. G., Melia, A. J., Searls, A. W. and Scarbrough, R. J. D., 'Semiconductor device developments: production and testing of highly reliable 4A-type transistors', *Post Office Elect. Engrs J.*, **61**, p. 7, 1968–9.

8. Ross, I. M., 'Reliability of components for communication satellites', *Bell Syst. Tech. J.*, **41**, p. 635, 1962.

9. Mayo, J. S., Mann, H., Witt, F. J., Peck, D. S., Gummel, H. K. and Brown, W. L., 'The command system malfunction of the *Telstar* satellite', *Bell Syst. Tech. J.*, **42**, p. 1631, 1963.

10. Back, R. E. G., Wilkinson, D. and Withers, D. J., 'Commercial satellite communication', *Proc. IEE*, **119**, p. 929, 1972.

11. Reeves, A. H., French Patent 852, 183 (1938); also 'The past, present and future of pcm', *IEEE Spectrum*, **2**, p. 58, May 1965.

12. Davis, C. G., 'An experimental pulse code modulation system for short-haul trunks', *Bell Syst. Tech. J.*, **41**, p. 1, 1962. Mayo, J. S., 'A bipolar repeater for pulse code modulation signals', *ibid.*, p. 25.

13. Bullington, K. and Fraser, J. M., 'Engineering aspects of TASI', *Bell Syst. Tech. J.*, **38**, p. 353, 1959. O'Neill, E. F., 'TASI', *Bell Labs Record*, **37**, p. 83, 1959.

14. Yokelson, B. J., Cagle, W. B. and Underwood, M. D., 'Semiconductor circuit design philosophy for the central control of an electronic switching system', *Bell Syst. Tech. J.*, **37**, p. 1125, 1958.

    Keister, W., Ketchledge, R. W. and Lovell, C. A., 'The Morris electronic telephone exchange', *Proc. IEE*, **107**, Part B, Suppl. 20, p. 257, 1960.

15. Nineteen papers on No. 1 ESS, *Bell Syst. Tech. J.*, **43**, pp. 1831–2592, 1964.

16. Long, R. C., 'TXE2 electronic switching system', IEEE International Conference on Communications, Philadelphia, 1968, p. 91.

17. Chapman, K. J. and Hughes, C. J., 'A field trial of an experimental pulse-code modulation tandem exchange', *Post Office Elect. Engrs J.*, **61**, p. 186, 1968–9.

18. Berry, R. W., Miller, P. and Rickert, R. M., 'A tone-generating integrated circuit', *Bell Labs Record*, **44**, p. 319, 1966.

19. Many aspects of this type of transmission system are dealt with in detail in the IEE Conference Publication No. 71, 'Trunk Telecommunications by Guided Waves', 1970.

20. Kao, K. C. and Hockham, G. A., 'Dielectric-fibre surface waveguides for optical frequencies', *Proc. IEE*, **113**, p. 1151, 1966. (An early paper on a possible system.)

    Kompfner, R., 'Optics at Bell Laboratories—optical communications', *Applied Optics*, **11**, p. 2412, 1972. (A current review.)

21. Crowell, M. H. and Labuda, E. F., 'The silicon diode array camera tube', *Bell Syst. Tech. J.*, **42**, p. 1481, 1969.

    Woolgar, A. J. and Bennett, C. J., Silicon-diode-array tubes and targets', *R. Telev. Soc. J.*, **13**, p. 53, 1970.

Manuscript received by the Institution on 24th November 1972. (Paper No. 1499/ Com 61.)

UDC 621.382:621.396.96

The impact of solid-state electronics on the development of radar systems

C. C. FIELDING, B.Sc.*

and

J. B. G. ROBERTS, B.Sc., Ph.D.*

SUMMARY

The paper indicates the status of radar at the date when the transistor was being developed and shows that the early impact of this device, because of its then limited performance, was primarily in the design of sub-systems to handle the large amount of data generated from radar receivers. In this respect the invention of the transistor is inseparable from the development of the computer.

During the '60s, however, the greater spectrum, together with improved performance, of semiconductor devices allowed for their penetration into all aspects of radar including in many instances the microwave power sources, to yield high performance, portability, reliability and lower cost per function.

The paper concludes with a survey of possible future links between solid state electronics and radar systems.

* *Royal Radar Establishment, St. Andrews Road, Great Malvern, Worcestershire WR14 3PS*

1 Background

Although use has been made of radio frequencies ranging from a few MHz to tens of GHz and of transmitted mean power levels from milliwatts to megawatts the layout of most radar systems conforms with that outlined in Fig. 1.[1,2]

A master timer triggers a transmitter to generate a pulsed waveform which is radiated from the aerial via the duplexer; this last device is essentially a switch which allows a common transmit/receive aerial by preventing direct transmission into the receiver. After transmission the duplexer switches the aerial to the receiver which consists of a radio frequency low-noise amplifier whose output is reduced to an intermediate frequency (usually a few tens of MHz) using a mixer. This generates an output at the difference frequency between the received signal and a local oscillator. There follows an i.f. amplifier and matched filter which maximizes the received signal energy with respect to the background noise level. The detector removes the i.f. carrier (e.g. by rectification and low-pass filtering) and passes on the modulation envelope, usually, to a display viewed by an operator. The normal display for radars with a search function is usually the plan position indicator (p.p.i.).

The schematic radar of Fig. 1 is fairly typical of the radars of the early 1950s, data being extracted manually by an operator and stored either on dials or as markers moved on a large-scale map in response to individual operator commands. The design was then based upon thermionic valves and the major effort was to seek high reliability and ruggedness from these devices in the light of increasing complexity in the design. In the majority of such simple systems a potentially valuable property of the echo, namely its phase information, was ignored. Measurement of each echo phase required so-called 'coherent' techniques which relied on the transmitter phase being remembered or known and then compared with the phases of received echoes. Echo phase changes from pulse to pulse are a manifestation of the well-known Doppler effect and can be used to separate moving targets from stationary clutter which can be 30–40 dB, or more, stronger. However, perhaps the biggest limitation was the inability to handle in some semi-automatic way the mass of signal data which was available from the simplest pulsed radars. This related both to signal processing and residual data processing where there was a need to be able to process and store wanted signal data over a period of many minutes in order to build up a clear geographical picture of the targets within the volume of coverage of the radar.

During the 1950s increasing use was made of phase information by the introduction of acoustic delay line filters to provide moving target indication (m.t.i.).[3] In its simplest form a phase-detected echo is compared with the equivalent echo from the previous transmission and a phase change indicates a moving target. Two or more delay lines were sometimes employed to improve moving target detection and stationary clutter rejection. Even more complete use of echo amplitude and phase information is provided by the 'pulse-Doppler radar': this is characterized by a series of range gates each one yielding a

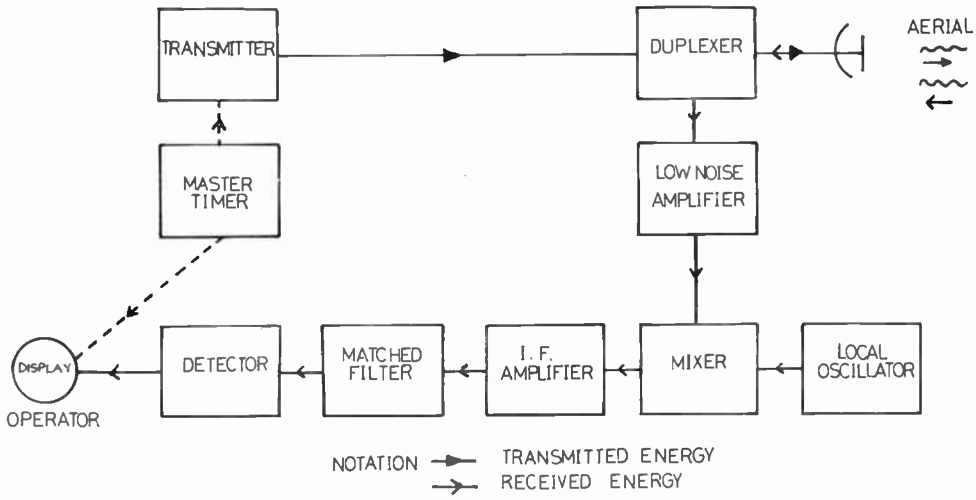


Fig. 1. Block diagram of a simple pulsed radar system.

Doppler frequency spectrum. Such a radar measures target range rate and under certain circumstances can detect a slowly moving target within the clutter frequency band by means of the target's spectral shape.

Attempts to carry out this kind of sophisticated processing using valves were not wholly satisfactory (they were usually based upon analogue systems) for the following reasons:

(a) resultant system unreliability;
(b) high power consumption and heat dissipation;
(c) high cost (not only in components but in the overall engineering framework);
(d) difficulty of achieving a satisfactory level of performance from an analogue system.

At this point in time valves did not represent a particular limitation in the evolution of the r.f./i.f. part of a radar system. The low-power receiving systems were not very complex and thus satisfactory designs could be produced by careful engineering. Lack of reliability was the chief limitation, mainly in the high power (transmitter) section of the radar.

The residual disadvantages of the thermionic valve-based radar may be catalogued as follows:

(1) lack of reliability;
(2) the need for large and unreliable power supplies;
(3) cooling and ventilation problems;
(4) heavy and bulky equipment;
(5) the difficulty of introducing any degree of automation in performance checking and fault finding.

2 The Early Impact of the Transistor

In the late '50s, semiconductor devices were becoming readily available and low in cost. The spectrum of devices was limited to p-n-p and later n-p-n discrete junction transistors together with low-power diodes, photo-diodes, etc. Although some devices were available at high cost operating up to some tens of MHz, the majority of low-cost devices had much lower cut-off frequencies (less than 10 MHz).

It is not surprising therefore that initial impact was on the low-frequency circuits in radar and perhaps most significantly grew from the rapid evolution of digital circuits and thence of digital computers.

Now for the first time, radar engineers could envisage the prospect of extracting and storing the geographical position of radar target data and carrying out subsidiary operations such as:

Establishing track, direction and speed; semi-automatic height-finding to give three-dimensional positional information.

Establishing and storing track identity.

Grouping of targets so that relevant information could be fed automatically to particular radar operators.

Allowing the operator to input information on tracks.

Making better use of manpower by freeing the operator from operations which could be better carried out by machines.

Coordinate conversion of positional data to provide super-imposition of track data from multiple radars at a central point.

For this purpose we can regard the evolution of the transistor as being closely coupled to the invention of digital computers. The major difference however between these computers and those which were developed for commercial work, was that they were required to operate in so-called 'real time', that is the processor would have to carry out its task in specific short time frames. An example of this might be the need to output all track data to each display position during a scan of the radar antenna.

As might have been expected, the military needs of radar were the incentive for innovation and a number of quite complex radar systems were conceived around this period. This applied particularly to ground-based air surveillance radars which had been developed to provide good resolution and ranges out to optical cut-off at heights beyond 60 000 ft. A radar of this type sited in the southern part of the UK will potentially have many hundreds of aircraft within its volume of coverage (see Fig. 2).

The benefits conveyed by transistor-based digital circuits meant that it was now possible to design a system to keep track continuously of all these targets. Not only that, but the information could now be stored in an orderly manner and thus be made available to particular parts of a system (whether it be local or remote) where reaction was necessary, in both defence and air traffic control. Such systems could be assembled within cost and reliability frameworks which would not have been conceivable in the valve era.

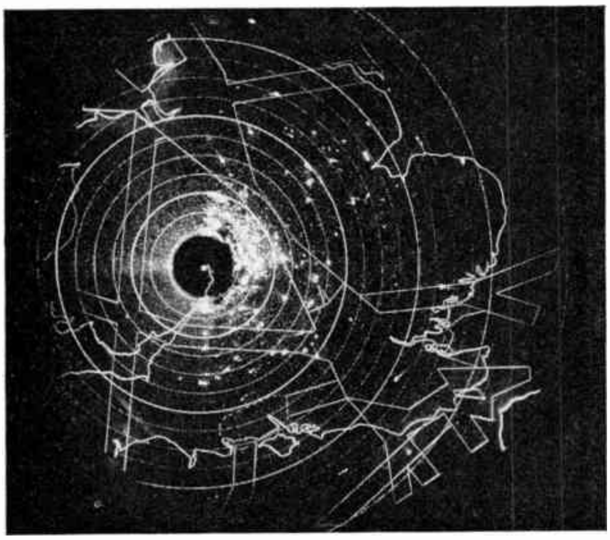


Fig. 2. P.p.i. display of aircraft and clutter echoes superimposed on a map of the British airlane system.

A corresponding advantage was to evolve from these computer-based systems, namely the ability automatically to monitor the performance of parts of a system and indicate fault conditions when they occurred. This was an essential step since such systems had by now become complex enough to demand the use of such automatic monitoring techniques in order to guarantee the performance of the equipment in the field.

3 Major Developments of the '60s

3.1 *General Trends*

This decade saw a number of significant advances in the development of radar, and the increasing spectrum of semiconductor devices played an important part.[4]

Perhaps the most significant factor has been the rapid evolution of silicon planar transistors towards more and more complex integrated circuits. Beyond this there has been continuous innovation of new microwave devices and circuits covering amplifiers (low noise parametric types and microwave transistors operating up to about 5 GHz), microwave oscillators, e.g. Gunn effect devices (using gallium arsenide), avalanche devices (using silicon and gallium arsenide), and frequency multipliers using varactor diodes. Hybrid microwave circuits could now couple solid-state active devices, e.g. local oscillators and mixers, on a compact transmission medium, analogous to printed circuits, called microstrip. Of somewhat less significance has been the invention of power switching devices of the thyristor class.[5,6]

Thus designers now had available a comprehensive range of compact, reliable circuits covering the frequency spectrum from radio frequencies (1 GHz–20 GHz) downwards. For many of the circuits, particularly those amenable to reasonable levels of integration, the cost per function had dropped dramatically by about two orders of magnitude. This stimulated a radar development pattern which can be divided into three general classes and these will be discussed in the following sections.

3.1.1 Large, complex radar systems

Although the underlying principles of radar, in terms of signal detection theory, were well understood and documented, designers had often been prevented from pursuing sophisticated designs by the unreliability of the resultant complex circuits. This situation was now substantially changed and complex solutions could now be developed, and without greatly increasing the cost compared with earlier designs.

The effort was still concentrated mainly on the data handling aspects of radar where complex computing systems were utilized to extract and store data, display it automatically to operators, carry out forward prediction of track information, and also the necessary calculations to provide control of aircraft, missiles, shells, as well as communications routing.[7]

However, the large amount of spurious data produced by most simple radars, as a result of echoes from clutter and weather, was still enough to saturate the data handling element (a long-range air surveillance radar will be required to examine more than 10^6 range resolution cells per scan). In the latter half of the decade, therefore, considerable effort was devoted to developing improved signal processing. This aspect has continued into the '70s as one of the residual problem areas in radar. It has been tackled, partly by improving the resolution of radars, primarily in the range dimension, and partly by using frequency discrimination to distinguish between returns from quasi-stationary objects and those from moving objects which will show a different Doppler beat frequency spectrum. As may be imagined, the resulting designs are very sophisticated and are only feasible using solid-state electronics. Figure 3 shows a computer generated display for air traffic control.

A further significant trend has been concerned with aerial design where increasing attention has been given to replacing the classical mechanically scanned systems with aerials in which the beam is electronically scanned in at least one dimension. Such systems work on the principle of distributing a large number of radiators over the aperture and controlling the phase of the signal to be transmitted from each radiator individually with respect to its neighbours. This approach necessarily relies on the availability of cheap reliable microwave components as well as the corresponding digital circuits which are required for the beam steering computer.

Cost is still a significant factor here and it has been left to the USA to implement most of the systems now in use. (See Fig. 4.)

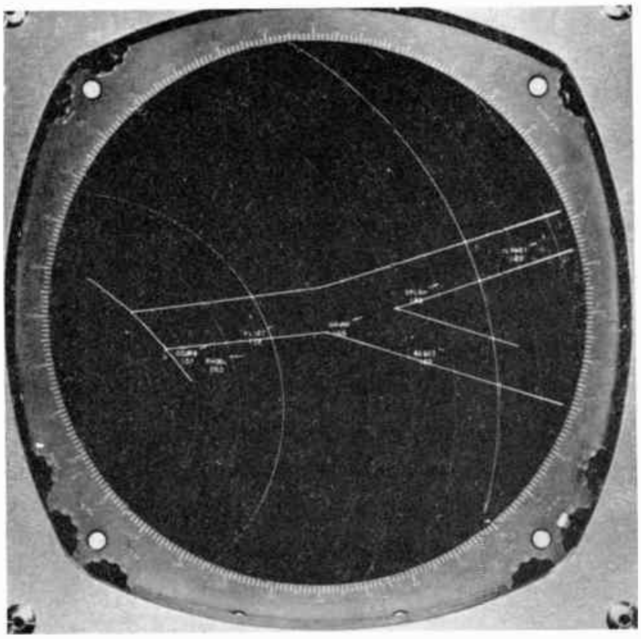


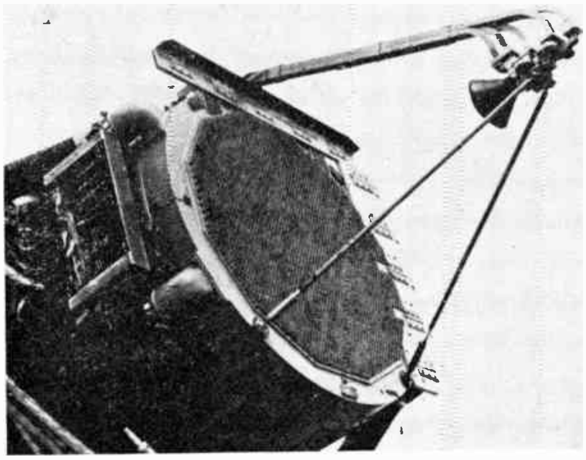
Fig. 3. A computer-generated p.p.i. display for controlling aircraft movements. Aircraft identity, position and height (relayed by transponder) are indicated as well as flight direction and speed in relation to an airway.

Utilization of such comprehensive systems extends from military air defence, through the important air traffic control field to the possible use of radars for weather reporting.[8]

3.1.2 Sophisticated mobile radars

The steadily increasing demand for mobile radars, particularly in the military field, has been partially thwarted in the past by the high bulk, relative unreliability and high power consumption of such systems. During the '60s, it was seen that the advent of semiconductor devices would allow highly compact systems to be developed, often with improved performance and requiring more modest power supplies. A striking example of this is shown in Fig. 5. The mobile mortar locating radar shown in Fig. 5(a) with its associated power generator totalled 7 tons in weight and has been replaced by a self-contained set in Fig. 5(b) which weighs less than one ton and has an improved radar performance.

This evolution has been mirrored in airborne systems where relatively crude, simple radars have now been progressively replaced by much more sophisticated, compact systems, usually with enhanced performance. The range of applications is quite broad, from essentially military applications like mapping radars to storm warning radars and radio altimeters.



(*Photograph by courtesy of the Raytheon Company*)

Fig. 4. The RARF airborne planar phased array radar intended to perform multiple interlaced functions. The array elements reflect the radiation from the horn feed with controlled phase to form a directed beam.

(a) Mortar locating radar with its power generator on a separate trailer.

(b) Modern counterpart of (a). The reduction in power requirements allows the generator to be incorporated in the single unit.

Fig. 5. Mobile radar development.

3.1.3 Mini-radars

This is the name which has been given to an extremely new class of radar application which stems directly from the innovation of solid-state microwave components. They are always low-power systems, usually 10–500 mW mean power and have already been developed for a large number of diverse applications. Of particular interest has been the hand-held radar (SPRAT), Fig. 6, which operates at 10 GHz, has a 12° beamwidth and allows the operator to detect targets by presenting to him the Doppler shifted returns on a pair of earphones. By this means, the presence of a walking man can be recognized at 1 km range.

Fig. 6. The Small Portable RAdar Torch (SPRAT) detects moving objects out to 1 km range by aural presentation of Doppler shifted echoes.

It should be noted that such radars are only practicable if they have small aerials without too great a sacrifice in azimuth resolution, i.e. they must work at very high frequencies. This has stimulated development of a range of solid-state microwave sources covering the frequency range 10 GHz to 40 GHz.

Other applications of such radars have been to aid the docking of large tankers, in which they measure the range and approach speed of different parts of the ship; to produce simple burglar alarms which detect even slow movements by their Doppler shifted returns; to make non-contacting measurements, e.g. shaft rotation speed measurement, again by Doppler frequency extraction; even the use of a small radar working at 35 GHz to examine the wing-beat frequency of a hover-fly which was too silent to be sensed acoustically.

3.2 *A Survey of Particular Devices*

It is worth while now tracing, in more detail, the influence on radar of devices which evolved during the decade in question.

3.2.1 Silicon integrated circuits

These allowed circuits to operate a hundred times faster than the early junction devices. This made possible powerful digital processors and hence greater target handling capacity. In addition, digital techniques penetrated into signal processing in cases where the radar bandwidth was small enough, for instance, discrimination against clutter became possible.

With the increasing complexity of radar systems it became essential to include self-checking facilities within the system design. This was now feasible.

Perhaps the chief impact of integrated circuits is epitomized by the fact that designers were now able to think in terms of large, complex systems, in which radar was one sensing element in an integrated information processor for, say, navigation or air traffic control.

3.2.2 High frequency transistors

Transistor amplifiers for receiver 'front ends' could be designed with good noise performances (a typical attainable noise figure being 6 dB) to operate up to 1 or 2 GHz. Oscillators working above 10 GHz were possible using varactor and step recovery diode frequency multiplier chains driven by relatively high power, low-frequency transistor oscillators stabilized by crystal control. These provided microwave sources with great frequency stability and low noise with sufficient power to act as local oscillators or to drive power amplifiers. Multiple stage i.f. amplifiers with complex gain characteristics offered great flexibility in signal processing options. Swept gain with a range of say 50 dB could be used to compensate variations in signal level due to range dependence and aerial illumination pattern, or a logarithmic characteristic could be provided to maintain a constant false alarm rate against certain types of clutter. Figure 7 contrasts the relative sizes of transistor amplifiers for 3 GHz (with and without an integral power supply) with a travelling-wave tube of comparable performance.

3.2.3 Hybrid microwave circuits

Thick and thin-film techniques for digital and analogue applications penetrated into the microwave field to provide very precisely designed and compact modules for low-power amplifiers, frequency multipliers, power amplifiers (up to several watts) and phase shifters. Microstrip circuit construction was particularly suitable for applications where many identical modules were required as for instance in providing phase shifters for a

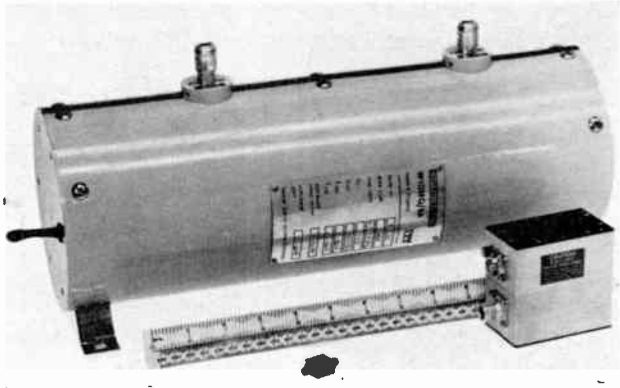
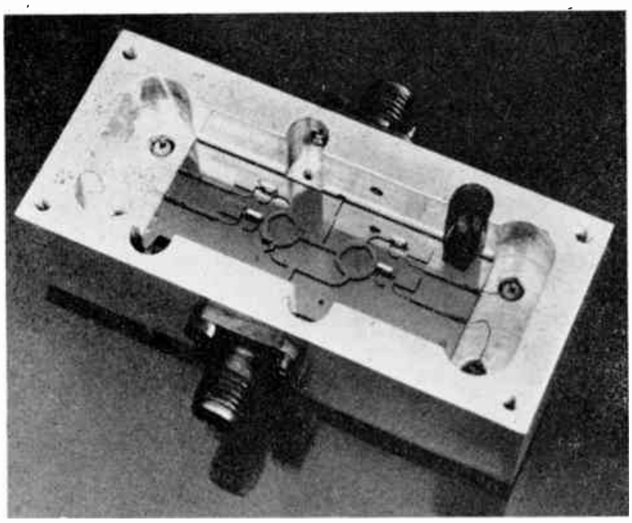


Fig. 7. A 3 GHz travelling-wave tube (t.w.t.) amplifier compared in size with a modern solid-state amplifier of comparable performance. Half the volume of the latter is occupied by the power supply whereas that for the t.w.t. is not shown. The actual amplifier is shown in the front of the picture.

phased array aerial. Figure 8 exemplifies a thin-film microstrip image rejection mixer.

3.2.4 Parametric amplifiers

A reduction in receiver noise level allows a lower transmitter power to be used to achieve a given radar sensitivity. A major contribution in this direction came from the type of parametric amplifier in which the capacitance of a varactor diode is 'pumped' at a high frequency, thus providing a negative resistance for the r.f. signal. The noise temperature of an uncooled 3 GHz parametric amplifier is typically 50 K (a noise figure of 0·7 dB); considerably better than a travelling wave tube and far more convenient than the maser which requires liquid helium temperatures. Satellite tracking represents an application where cooling a 'paramp' to liquid nitrogen temperatures is justified to gain a further reduction in noise temperature.



(Photograph by courtesy of GEC Limited)

Fig. 8. A hybrid circuit image rejection mixer using 4 l.i.d. diodes (each with an r.f. rejection filter in the output line), two hybrid couplers, two power dividers and an impedance matching transponder.

Figure 9 shows a broadband parametric amplifier with 20% bandwidth at 4 GHz which provides 10 dB gain. Its noise temperature when operated without cooling is about 100 K.

3.2.5 Solid-state microwave oscillators

The discovery of the Gunn diode in 1963 ushered in the era of solid-state microwave oscillators. The Gunn diode, made from the semiconductor material gallium arsenide, operates at a frequency determined by the length of the device; once each cycle a 'domain' transits the device. By 1970 Gunn diodes had been demonstrated as sources in burglar alarms, in miniature radar systems and as local oscillators.

In 1966, a new mode of operation in gallium arsenide was discovered, called the l.s.a. mode. By overcoming the restrictions whereby the length of the device determines the operating frequency, much greater output powers were obtained. Peak powers of many hundreds of watts

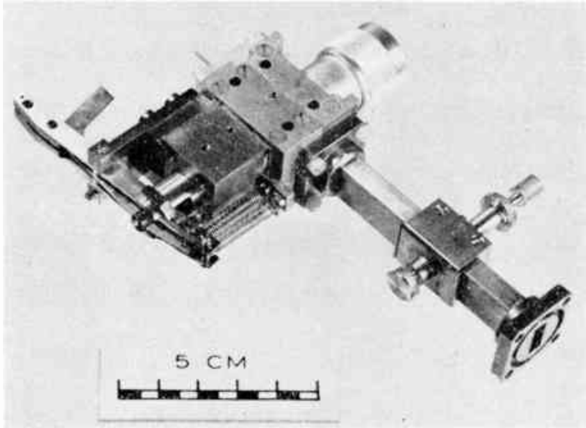


Fig. 9. A 10 dB gain parametric amplifier for 4 GHz which provides a response with less than 0·33 dB gain variation over 770 MHz.

at 5 GHz were achieved when driving the device with a pulsed bias supply.

Paralleling these developments was the development of avalanche diodes as microwave sources. These devices are semiconductor p-n junctions, in silicon or gallium arsenide, which are driven into avalanche by a reverse bias voltage. The first oscillator was made in 1964, and the development from this discovery in the impatt diode. In 1967 a high-efficiency mode of operation was discovered which results in conversion efficiencies of up to 60%. Devices which work in this high efficiency mode are known as trapatt diodes.

Figure 10 shows the convenient size of some 10 GHz sources made in 1966 which provide 30 mW of c.w. power at 4% efficiency from a 9 V supply. The complete cavity assembly weighs less than 5 grams.[9]

3.2.6 Thyristor switches

This power switch uses four layers of material in p-n-p-n sequence. Substantial powers can be switched quickly (approximately 1 μs) from low power triggers. Radar systems benefited in two areas: pulse modulators in which a high power video pulse is used to switch on a microwave power oscillator such as a magnetron, and

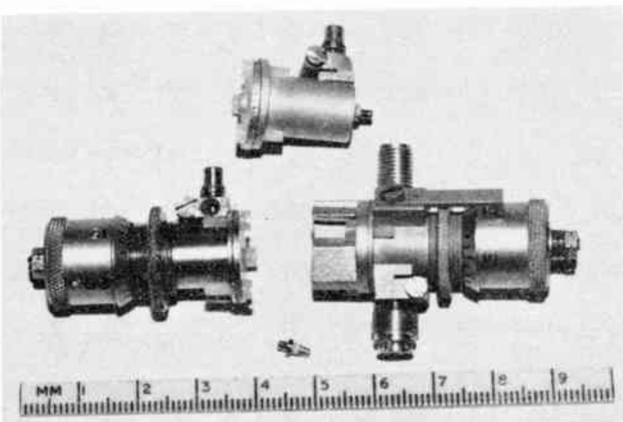


Fig. 10. Gunn oscillators for 10 GHz operation: (upper) fixed frequency type; (left) mechanically tunable cavity with 2 GHz range; (right) mechanically and electrically tunable version; (lower) the mounted Gunn diode.

inverters used for efficient d.c. to d.c. or d.c. to a.c. voltage conversion. The latter is important in the design of flexible, efficient power supply systems.

3.2.7 p-i-n diodes

These provide convenient limiters and switches to handle microwave power up to and beyond 1 MW peak. Limiters of this kind are required as part of the duplexing function of a radar in which the aim is to prevent damage and saturation in the low-power receiving circuits when a common aerial is used for transmission and reception. Duplexers and phase shifters frequently depend on p-i-n diode switches which are controlled by application of a d.c. bias.

3.2.8 Mixers

The mixer diode has been subject to continuous development since the early days of radar. Point contact designs continued in use and germanium devices were developed with noise figures of 6 dB, including an i.f. contribution of 1·5 dB. Much development was also directed towards establishing Schottky-barrier devices produced using a planar process. The work concentrated on silicon and gallium arsenide devices, the latter giving noise figures of 5·5 dB in the frequency range 1–10 GHz.

4 Prospects for the Future

The rising demands of military and civil radar users ensure that the techniques make full use of advances in solid-state electronics.

4.1 *Clutter Rejection*

One of the prime areas where better performance is required is in removing clutter from displays and particularly from radar data intended for computer processing. The intention here is to enable smaller targets to be detected against strongly scattering backgrounds (e.g. a battlefield or sea surface) and to avoid overloading an automatic processor with unwanted detections. The chief weapons against clutter are good spatial resolution (involving high bandwidth transmissions say > 20 MHz) and moving target indicator (m.t.i.) techniques which may use coherent Doppler frequency filtering or reject apparent targets whose position does not change with time (area m.t.i.).

The combining of high bandwidth and reasonably high duty cycle (avoiding a very high peak power transmission) is becoming far simpler with the appearance of pulse compression systems based on surface acoustic waves. These use piezo-electric coupling to launch and receive elastic waves along the surface of a crystal (e.g. quartz). The low acoustic velocity ($\sim 10^5$ cm/s) enables delays of many microseconds to be obtained, with frequency dispersion characteristics accurately defined by the layout of the transducers. Fractional gigahertz bandwidths are anticipated from these convenient passive components.[10]

Coherent m.t.i. is moving from the era of bulk delay line cancellers to digital filtering which offers great precision in design and increasing bandwidth capability as digital processing speeds improve. More novel kinds of delay line in the form of analogue shift registers are now being developed for signal processing and it is possible that these will perform better than fully analogue or fully digital delay lines. These devices, known as charge-coupled or 'bucket brigade' circuits, transfer sampled analogue signals along a chain of capacitors and potentially offer the precise but flexible control of the time variable attainable in digital processors without the complexity of high speed digital multipliers.

Area m.t.i. may become important as mass digital storage becomes cheaper and faster. The range-azimuth co-ordinates and probably amplitudes of detections have to be stored for comparisons with succeeding scans and therefore good spatial resolution involves storing a very large amount of information for serial access in the case of a uniformly scanning radar.

4.2 *High Resolution and Accuracy*

Accurate ranging, measurement of bearing and radial velocity demand careful control of the transmitted waveform and sophistication in the processing of the returned signals. Surface acoustic wave devices have much to contribute to both of these by allowing a high mean power (and therefore good signal/noise ratio) to be combined with fine range discrimination.

With progress in large scale circuit integration, digital signal processing may be expected to advance further. Digital filters, already in use for m.t.i., may be used for real-time full Doppler spectrum analysis, perhaps employing the fast Fourier transform algorithm which has enormously speeded-up computer frequency analysis. Digital data processing will no doubt make great strides also, using stored information to refine the accuracy, reliability and the interpretation of radar displays. As digital pattern recognition techniques improve, it may be possible to classify targets, given extremely good resolution and a high signal/noise ratio.

The technique of aperture synthesis, whereby signals received at several points in space are combined to achieve an azimuth resolution related to the dispersion of the receiving points rather than to the real aerial aperture, has received much attention, particularly for terrain mapping by airborne radar. There is every possibility of digital processing becoming adequate to derive the high resolution map in real time aboard the aircraft within the next few years.[11]

4.3 *High Integrity Radar Systems*

The extreme reliability of solid-state electronics, particularly in integrated form, will encourage the development of self-contained systems intended to perform without attention for long periods in adverse environments, for instance in military aircraft or at remote unattended sites relaying meteorological or other data to an information centre. The reliability of the processing system will be as important as that of the radar itself and it will need to be capable of working in a range of circumstances encompassing variations of clutter severity (due to weather conditions perhaps) and target density. Ways of controlling the false alarm rate are likely to be an important feature of such systems.

4.4 *Multi-function Radars*

Many large and small radar systems will be required to perform more than one simultaneous task, for instance automatically to track several aircraft (returning to check their positions as needed) while surveying space to ensure that new targets will be acquired by the system. Such a system would be computer based and might carry information processing one or two stages further than is normal at present. For instance, targets could be interrogated, using transponders, for status information, and potential collision warnings issued. With such flexibility, automatic and self-adaptive control will be needed. The prime essential before these systems can come into operational use is the development of cheap reliable phased-array aerials. These in turn are dependent upon the development of cheap, mass-produced phase shifters and rapid progress is being made here.[12]

4.5 *Small Special-purpose Radars*

Low-voltage compact solid-state microwave sources have already found significant use in 'mini-radars'. It seems likely that this almost qualitatively distinct field of radar applications will continue to expand. Quasi-continuous wave rather than pulsed transmissions will often be used, with frequency modulation to allow range information to be extracted. Foreseeable roles will include the control of vehicle spacing on motorways, counting moving objects, measuring speed, burglar detection and general area security systems.

4.6 *Radar Displays*

A major component of most radar systems is the display console. Presently-conceived systems rely on the cathode-ray tube which has been progressively developed into a high definition, relatively reliable component. In addition, solid-state electronics has contributed to the feasibility of developing complex drive systems for c.r.t.s allowing sophisticated time-multiplexed displays to be produced. Even high brightness systems can be produced by this means. Where there is a requirement for displaying quantities of data at high resolution therefore, the current systems will prove very difficult to dislodge.

However, solid-state display elements using both electro-luminescence and liquid crystal devices are being actively developed and these together with miniature gas discharge tube assemblies will make an impact over the next decade; attacking first of all the requirement for quasi-static indicators and following with low resolution displays, e.g. alpha-numeric lists, etc.

5 Conclusion

Radar has improved very considerably over the past four decades and, in analysing the technical advances which have contributed to this, it is clear that they have come from a mixture of good techniques innovation together with the advent of new components, particularly those based on semiconductors. Among those improvements attributable to the availability of semiconductor devices, it is possible to single out the rapid development which took place during the '60s of sophisticated radar data handling systems based on digital circuits. This was made possible by the ready availability of cheap, reliable silicon circuits based on planar technology.

Although the essential principles and limitations of radar are well understood, and this implies that we can expect no dramatic improvements in radar performance, the continuing substantial developments in semiconductor technology will continue to make orders of magnitude improvements to the complexity, reliability, cost, size and power requirements of the various system components.

The cumulative result of piecemeal improvements in inertialess scanning, r.f. hardware, signal processing, data handling and displays may well be to open up new fields of application for radar in the same way that the emergence of small high frequency microwave generators has revealed a new market sector for mini-radars.

6 Acknowledgments

The authors are indebted to a number of their colleagues for providing them with background material for this paper and this is gratefully acknowledged. Further, acknowledgment is made to the Director of the Royal Radar Establishment for permission to publish this paper.

7 References

1. Skolnik, M. I., 'Introduction to Radar Systems', (McGraw-Hill, New York, 1962).
2. Cook, C. E. and Bernfeld, M., 'Radar Signals—an Introduction to Theory and Applications', (Academic Press, New York, 1967).
3. Barton, D. K., 'Radar Systems Analysis', (Prentice-Hall, Englewood Cliffs, N.J., 1964).
4. Fitchen, F. C., 'Electronic Integrated Circuits and Systems', (Van Nostrand Reinhold, New York, 1970).
5. Shurmer, H. V., 'Microwave Semiconductor Devices' (Pitman, London, 1971).
6. Watson, H. A., 'Microwave Semiconductor Devices and their Circuit Applications', (McGraw-Hill, New York, 1969).
7. IEE Conference on Electronics Research and Development for Civil Aviation, 1963.
8. Skolnik, M. I., 'Radar Handbook', (McGraw-Hill, New York, 1970).
9. Bulman, P. J., Hobson, G. S. and Taylor, B. C., 'Transferred Electron Devices', (Academic Press, New York, 1972).
10. Kino, G. S. and Matthews, H., 'Signal processing in acoustic surface wave devices', *IEEE Spectrum*, 8, pp. 22–5, August, 1971.
11. Hager, R. O., 'Synthetic Aperture Radar Systems—Theory and Design', (Academic Press, New York, 1970).
12. Oliver, A. A. and Kittel, G. H., 'Phased Array Antennas' (Artech House, Dedham, Mass., 1972).

Manuscript received by the Institution on 7th November 1972.
(Paper No. 1500/AMMS 54.)

UDC 621.382.3:681.32

The influence of semiconductor devices on the evolution of computer systems

G. G. SCARROTT, C.Eng., M.I.E.E.*

SUMMARY

The shortcomings of mechanical and vacuum tube computers are pointed out. The semiconductor device has provided greatly increased intrinsic reliability but has made apparent the problem of manageability. The semiconductor approach allows the classical methods of tackling complexity by partitioning and decentralization to be followed. Storage management and store searching aids can easily be realized by additional circuits.

* *International Computers Ltd., Research and Advanced Development Organization, Argyle Way, Stevenage, Hertfordshire SG1 2BD.*

1 Introduction

All present-day computer installations rely heavily on the use of semiconductor electronic devices, and this situation has applied for at least the last decade. At first sight, it would therefore appear that semiconductor devices have done much more than influence computer systems, they have made them possible. Such a judgment would, however, be an over-simplification since information systems existed long before semiconductor technology. Since excellent papers are already available describing semiconductor logical techniques, this paper will be an assessment, based on the historical record, of the effect of the introduction of semiconductor devices against a background of the longer term evolution of information systems, and will lead to some reflections on the direction which computers may take in response to the new technological situation created by the introduction of large-scale semiconductor logical systems fabricated by powerful microengineering techniques now coming into use.

2 What is the Present Situation in Information System Design, and How Did We Get There?

2.1 *Brief History of Device Foundations of Information Systems*

To get a clearer picture of the problems and opportunities which have been generated by the introduction of semiconductor devices into information systems, it is necessary to sketch in the main features of the history of the device foundations of artificial information systems.

For the purpose of this discussion three stages in the evolution of computer systems to their present form can be recognized, based on the use of mechanical devices, vacuum tubes and solid state devices respectively. During the first two of these stages the manufacturing problems pushed the available fabrication techniques up to, and sometimes beyond, the limits of their practicability.

Babbage's machine, based on the use of gear wheels as logical devices, was never finished on account of the inadequacy of the fabrication techniques available to him, and some of the early vacuum tube machines had to be treated with extreme care to ensure that there was a reasonable chance that they would not break down before a useful result was obtained. Now that semiconductor machines are with us, the manufacturing and reliability problems have moved into the background, and have been replaced by a host of new problems such as those concerned with keeping software development and maintenance costs under control and ensuring that the information processing power is not squandered on excessive system administration overheads. It was to be expected that the solution of one set of problems would reveal others, but the intractability of these system organization problems has taken some computer designers by surprise.

2.2 *An Interpretation of the History*

Although the facts of the history outlined above are probably widely accepted, the following interpretation of them is necessarily personal.

One may assume that information systems, like other engineering products, evolve in a way which enhances their 'usefulness' to society. We therefore need to have an understanding of what an information system is, and what determines its usefulness.

We may regard an information system as an assembly of logical and storage devices, together with programs for controlling it. The 'usefulness' of such a system depends on its size, measured crudely by the number of elements in it, and its manageability, determined by the logical structure of the assembly, and of the programs. However, the problem of manageability does not arise unless the size is above a critical threshold.

The practicable size in turn depends on the intrinsic reliability of the logical devices of which the assembly is composed. The reliability of a device may be measured by the probability of failure per operation. This measure can be used directly to assess the number of such devices which can be assembled into a system still leaving an adequate chance that the whole assembly can carry out a number of operations sufficient to tackle a problem of practical interest.

A very good mechanical device, e.g. a watch escapement, can carry out about 10^9 operations without failure, whereas a vacuum tube device may typically carry out 10^{14} operations, and a solid state device 10^{16} operations.

The low reliability figure of mechanical devices shows that only a small assembly of mechanical devices is practicable, so small that the problem of making it manageable hardly arises. Indeed, few men have difficulty using a mechanical desk calculator. Even when vacuum tube computers were introduced, for a long time the main problem was still hardware reliability, so that the problem of manageability which was lurking in the background was still not visible.

Now that semiconductor devices, with their greatly increased intrinsic reliability, have been introduced the problem of manageability is plain for all to see. For example it is notorious that large software packages tend to be inefficient and costly to maintain. Some users complain that 'the computer never does what I want it to do at the time when I can specify my wants'. Other users have noticed that if measures are taken, e.g. by the provision of multiple consoles, to make efficient use of human resources, then the machine is inefficiently used, and vice versa.

All these, however, are different aspects of the same central problem of partitioning complexity in such a way as to make it manageable by ordinary men. This problem has arisen following the introduction of semiconductor devices simply because they are so reliable that we can now make an assembly of hardware and software above the critical size, so that unless deep and careful thought is put into the design of the structure of the system, it often turns out to be difficult to use in one or more of the senses referred to above.

2.3 *The Present Situation*

To describe the present situation in system design it is useful to contrast the design objectives for the detailed components of an information system, e.g. storage cells, gates, individual instructions, with the design objectives for the large-scale operations such as an operating system or a compiler. In the former case it is possible and necessary that the design be absolutely right in the strict sense that an algebraic identity is true, so that the design problem can be regarded as an exact science.

The larger scale structures and operations, however, are inevitably complex, wasteful, and frequently modified so that the problem of their creation, maintenance, and use is properly concerned with minimizing the nuisance value of inevitable errors and reducing their number but not primarily with attempts to eliminate them altogether. Thus the techniques appropriate for the design of the large-scale structures can be more usefully borrowed from the humanities than from the exact sciences.

Present information systems have evolved largely by the insidious accumulation of high-level complexity. Designers are rightly influenced by their experience but some have allowed themselves to be trapped by it. They have been dazzled by the apparent versatility of their processing devices and preoccupied with their detailed operation and in consequence have paid too little attention to the principles which should be followed in the design and use of complex organizations to ensure that they are manageable.

3 What is the Problem of Complexity and How Can It be Solved?

The complexity problem is a feature of large-scale human activities and organizations. It arises in an elementary form whenever a human project is too big to be undertaken by one man, and becomes especially significant whenever high individual skills must be exercised by a large number of men cooperating in a team. The problem then arises of structuring the project and the teams of men who design and use it, so that it is a coordinated whole, even though no individual can take total responsibility for all the design compromises. Evidently there are many solutions to any particular manifestation of the complexity problem, so that the discovery of a solution is more an art than a science.

The complexity problem in the guise which has arisen in the design and use of information systems, appears to those who suffer from it to be an entirely new phenomenon. However, it is really as old as the pyramids, and so common human experience is a valid source of inspiration for finding solutions. For the purpose of this discussion three items of classical wisdom for solving the complexity problem in human organizations can be selected, which are particularly relevant to tackling the problem in information systems. They are:

(1) The problem cannot be solved by the refinement of artisan skills, e.g. no amount of weapon training will turn a mob into an army, a command structure is also necessary.

(2) Partition the problem by specifying interfaces.

(3) Decentralize.

Each of these principles can be applied equally to the design of the product and the design of the design teams, and it will be shown that their application to information systems can be facilitated by the same powerful semiconductor techniques which have revealed the complexity problems in the first place.

4 How can Semiconductor Technology Help the Solution of the Complexity Problem?

The superior reliability and low cost of semiconductor devices has made it no longer necessary, or even desirable, for system designers to seek too diligently the smallest collection of logical hardware which can do the job. It is well known in ordinary human affairs that some redundant facilities in a smoothly working human organization are inevitable. They arise because the partitioning of the main activity into comprehensible and therefore workable large-scale parts following the principles outlined above necessarily involves some duplication of the lower scale activities. Although at first sight such duplication often offends the pathologically tidy mind, experience has shown that excessive endeavours to avoid low-level inefficiency often lead to more costly high-level waste (i.e., 'penny wise, pound foolish').

It is therefore fortunate that the very reliability of semiconductor devices which has revealed the problem of complexity, also makes it possible to use the classic methods of solving it. Thus if the design of an information system is partitioned following the principles outlined above there will inevitably be a set of separate teams undertaking various well-defined parts of the design who will propose solutions to the problems that have been put to them. Each such solution may necessitate a hardware logical assembly, and each design team can be given the freedom to propose such a low-level autonomous subunit on account of the low cost and high reliability of semiconductor logical networks, even if their collected proposals involve some duplication of low-level processing facilities.

The first principle quoted in Section 3 when applied to information systems simply asserts that the complexity problems of computer system design are unlikely to be solved solely by the diligent application of methods already in wide use. This is a hard thing to say but its truth will be recognized by many practising computer engineers.

The other principles, that the problem should be systematically partitioned and the components of the problem sub-contracted to simpler sub-assemblies necessitates the recognition of tasks which must be carried out by the information system so frequently that it is right to devise recognizable sub-units to carry them out. An obvious candidate for this is storage management.

4.1 *Store Management Aids*

A typical information system includes several physical storage devices with different speeds and different storage capacities.

Access to information is normally made to an explicit address. Information so accessed can be used to specify or calculate a new explicit address. A chain of such explicit accesses constitutes a 'navigable path' which leads to the item of data ultimately required, via intermediate items of information which may be called 'pointers'.

When several programs or processors operate on the same information the problem arises that an error in one program may cause the pointer information to be disturbed and thus sabotage other programs. Attempts to prevent this consequential failure mechanism with the aid of a paging system have been only partially successful and it has long been recognized that to provide the simplest and most error proof storage protection necessitates the use of a few extra descriptor bits per store word which distinguish pointers from data.

Up to the present the inflexibility of ferrite core stores combined with the conservatism of designers has discouraged the use of such organizational devices but now that semiconductor stores have arrived it is a trivial matter to provide the extra bits and the equipment for handling them.

This single measure will lead to a substantial simplification of software and the software development process.

4.2 *Store Searching Aids*

Sometimes the navigable path to information is so complicated that it is impractical to keep the navigation data (the 'map') up to date, and it is then necessary to search the stored information systematically.

In present computer systems the searching operation, where it is unavoidable, is typically implemented by software which searches by exhaustive navigation, a clumsy and inefficient process. In consequence a natural selection process has tended to discourage computer applications which necessitate excessive data access by searching, so that we may reasonably assume that there is a pent-up demand for computer applications which necessitate more efficient store searching methods.

With the spread of time-sharing multi-access systems a new problem has arisen, the continuous allocation of system resources, e.g. stores or peripheral devices, to a floating population of computing jobs, a task which naturally includes a significant amount of store searching for unallocated resources. This gives emphasis to the need for more efficient store searching methods.

This requirement can easily be met by semiconductor technology since it is quite practicable to add to each random access storage chip a counter and comparison device such that the information can be accessed either from an explicit address as a step in the navigation process, or by a slower autonomous associative search for those cases for which the navigation process is impractical.

Moreover if it should be found that significant cost savings can be achieved in semiconductor stores by organizing them in the form of shift registers the resultant cyclic access can be turned to good account by the inclusion of a comparison device on each chip to provide associative access to the stored information.

Although reasons as above can be given for adding an autonomous search facility to the traditional explicitly addressed stores, no case can be made for omitting the explicit address mechanism.

4.3 *Information Retrieval*

The essence of the information retrieval problem is that on account of the intrinsic unpredictability of human affairs, the criteria which will be used to select information for retrieval are unknowable at the time the information is stored. It follows that the 'map' needed to navigate a path to the required item of information cannot be complete, although in practice useful parts of it may often be available. A solution to the information retrieval problem therefore normally involves a combination of navigation and search since the full navigation process is not only impractical, but impossible.

Evidently the store management and searching aids made possible by semiconductor technology are likely to be powerful tools for solving the information retrieval problem.

4.4 *Very Powerful Processors ('Number Crunching')*

In present practice it is not very useful to try to make very powerful processors in the form of a cooperating team of less powerful processors since the store management process is so intricately tangled with the program that attempts to make several processors cooperate on the same task usually lead to excessive administrative overheads. However, by the use of the store management and searching aids outlined in 4.1 and 4.2 the store management problem can be largely sub-contracted, so that in such a system it is quite practicable to use several independent processors operating on the same problem since their mutual interference is less than in present practice.

4.5 *The Use of Dynamic Microprogram Stores*

With the availability of low-cost fast storage, it has recently become practicable to make the hardware processor much simpler by putting most, but not all, of the control information into a fast store, known as the microprogram store. This technique leads to substantial reductions in both the hardware cost and the hardware development cost and moreover provides valuable low-level flexibility which in combination with the store management aids shows good promise of achieving a unique combination of efficiency and software simplicity.

4.6 *Pattern Recognition*

A potential field for the application of information systems is the recognition of patterns of such subtlety that at present the work has to be done by the use of skilled manpower (e.g. in preventive medicine).

It is likely that a solution to this problem will comprise a combination of a new peripheral device incorporating logical hardware to extract simple features from the pattern, together with complex and flexible adaptive software which leaves the operator the freedom to specify by example the patterns of interest.

The introduction of semiconductor technology will no doubt be of great value for building the new peripheral devices incorporating elementary feature detectors.

5 Conclusions

The conceptual foundations of the computer industry were laid by mathematicians who deduced that there was no limit to the potential of a perfectly reliable simple automaton obeying an arbitrarily complex program, but they did not look very closely into the problem of creating such a program.

Only by the introduction of semiconductor devices were engineers able to build an automaton sufficiently reliable to put these theories to the test and in the event the pioneers were proved right on their first point but had their bluff called on the second. In consequence it is now necessary to redefine the automaton, without seriously damaging its simplicity, to include facilities which can localize the effect of software errors. Fortunately the implementation of such a redefined automaton can be greatly assisted with the aid of semiconductor techniques now coming into use.

Manuscript received by the Institution on 21st September 1972. (Paper No. 1501/Comp 143)

UDC 628.382 : 621.397.62

Transistor circuits in television: some evolutionary aspects

A. MARTINEZ,
C.Eng., F.I.E.R.E.*

SUMMARY

This paper describes some aspects of transistor circuit developments in television receivers in the UK. Improvements in transistor specifications and cost, as well as in the conceptual approach to their use in television circuits are outlined briefly. Some of the circuits of one of the first fully transistored receivers produced in this country are compared with two colour television receivers in current production.

* *Baird Television, British Radio Corporation (Bradford), Ltd., Lidget Green, Bradford 7, Yorkshire.*

1 Introduction

The invention of the transistor provided a magnificent new brick for the circuit architect. Advancements in semiconductor technology can now offer a variety of building materials which in television receivers have given forth improvements in performance, stability, reliability, cost and compactness.

The introduction of transistor circuits has been influenced all along by transistor performance and cost. The fully transistored television receiver became a reality when transistors, capable of scanning the line circuits, were made available at economic prices.

The first transistor circuits to be developed for television were small-signal circuits and audio output stages, similar to those employed in portable radio receivers. Germanium devices were the first used and the stabilization of their quiescent operating conditions versus temperature imposed limitations in the maximum operating conditions. These limitations were overcome, in certain cases, by means of thermistors but seldom did this approach represent an overall solution. The advent of the silicon transistors, with their superior thermal characteristics, was perhaps the most significant step in the process of transistorization in television.

In a short paper one cannot hope to be exhaustive, let alone convey the impact that semiconductor technology has made in the television industry. In an effort to show some of the progress which has been made in the last decade, some of the circuits in the Perdio *Portarama* have been chosen as being representative not only of one of the first transistored television receivers manufactured in this country (in 1962) but because these show transistors being used in very basic forms often as direct replacements to valves.[1]

By contrast, and as comparison, some circuits of the British Radio Corporation 3500 and 8500 chassis are included which represent the more sophisticated and much more complex utilization of transistors found to-day in television circuits.[2, 3]

2 The Tuner

The improvements in the high-frequency capability of the transistor is best illustrated by looking at the front end of the receiver. The *Portarama* operated at v.h.f. and transistors with cut-off frequencies high enough to obtain the required gain at Band III were available when this receiver was manufactured.

Very shortly afterwards, with the introduction in the UK of the 625-lines system, u.h.f. operation had to be achieved. Both the transistor and the tuner manufacturers rose to the occasion and operation at the top end of Band V was achieved after an initial struggle. Further improvements in cut-off frequencies and in the noise performance of these transistors have taken place since these early days and noise figures of 9 dB for the top end of Band V are not uncommon to-day.

The first transistored v.h.f. tuners used a pre-set rotary mechanism for station selection. For u.h.f. variable air-spaced capacitors became the norm.

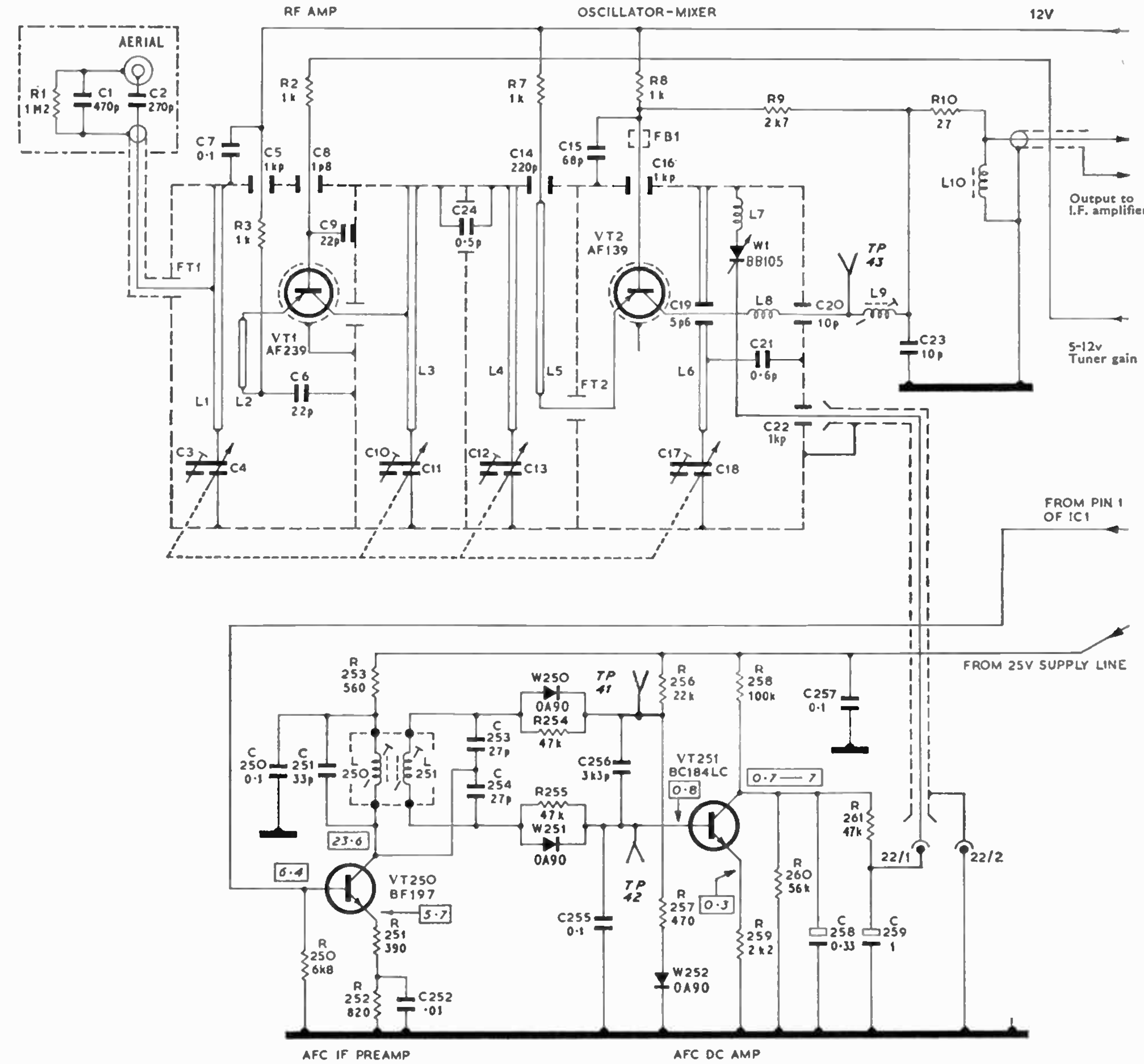


Fig. 1. B.R.C. 8500 series mechanical tuner, with varicap diode and a.f.c. circuit.

The development of the varicap diode enabled the circuit engineer to improve the performance of mechanical tuners first by the addition of automatic frequency control. The need for mechanical tuning was then eliminated by the introduction of the varicap diode tuner. The varicap diode has become the leader of the front-end 'revolution' offering improvements in stability, long-term reliability and versatility, developments into the methods of station selection with 'touch tuning' and further simplifications to remote control station selection.

Figure 1 shows a typical mechanical tuner with varicap diode for a.f.c. operation together with its control circuit as employed in the 8500 chassis.

In Fig. 2 the circuit diagram of a basic varicap tuner fitted to some of the models employing the 3500 chassis can be seen.

3 The I.F. Amplifier and its Gain Control

In television, i.f. amplifiers were developed first, using germanium transistors operating very close to the limit of their high frequency capability. Their collector-base feedback impedances were such that neutralization had to be employed; alternatively, the gain per stage was kept low to avoid the need for neutralization and extra stages added in order to obtain the overall gain required. Margins of stability were low and consistent gain-bandwidth specifications were difficult to achieve.

For early receivers automatic gain control presented difficulties which were not overcome until forward bias transistors with suitable characteristics became available. At first, in an effort to solve this problem, reverse bias characteristics were employed which gave rise to cross modulation. Forward bias diode circuits were used to

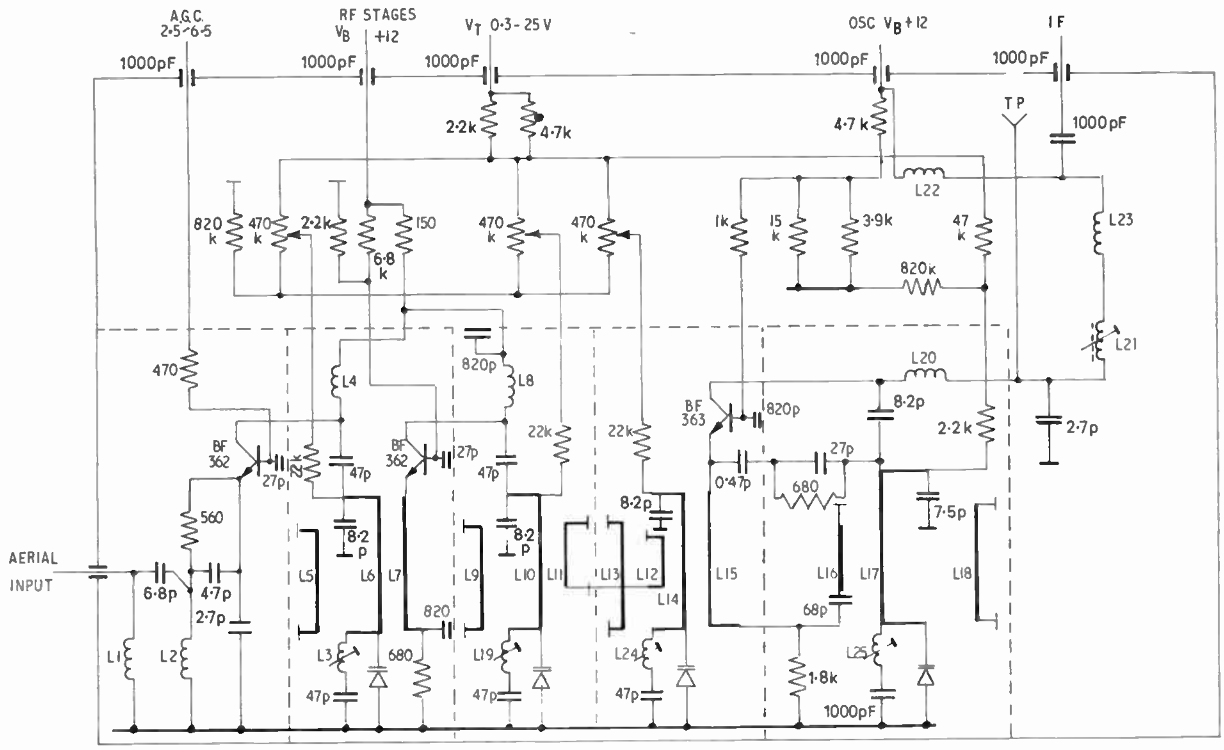


Fig. 2. Varicap tuner as used with 3500 chassis.

damp tuned circuits and thus decrease the gain, in so doing causing detuning and bandwidth changes. In some instances the collector-emitter voltage was lowered in order to decrease the cut-off frequency and thus decrease the gain.

Developments in transistored a.g.c. circuits followed the general pattern of the valve counterparts, namely mean level and gated a.g.c. employed in 405 system receivers, and peak sync rectification systems with 625 lines and negative video modulation. The availability of p-n-p and n-p-n combinations tends to simplify the transistor versions of these circuits.

4 The Power Supply

The three receivers chosen show considerable differences in the method of generating their power requirements from the mains supply. In each case the regulator employed satisfies the requirements defined by the voltage capability of the line output transistor. This meant a low power supply voltage for the *Portarama* (11·5 V), medium for the 3500 chassis (65 V) and high for the 8500 series (180 V).

The Perdio *Portarama* was designed as a mains battery portable with facilities to re-charge its own batteries. (For simplicity the switching facilities for battery operation and re-charging are not shown.) The basic circuit (Fig. 3) consists of a fully-isolated mains transformer with a secondary winding driving a full-wave rectifier and a reservoir capacitor. This is followed by a series regulator and a simple control circuit.

The power supply of the 3500 series chassis is shown in Fig. 4. The main supply consists of a series regulator employing chopper stabilization. With the power requirements of this chassis the provision of a step-down transformer would have proved both bulky and uneconomical. The power required at 65 V made this type of stabilization ideal. The chopper frequency was chosen to be synchronous with the horizontal deflexion system, in so doing power is pumped into the load at the same rate at which the majority of this power is dissipated. Regulation is achieved by changing the on/off conduction period of the series regulator.

Auxiliary supplies include a base stabilized emitter follower to generate 30 V, a u.h.f. tuner supply and a high voltage video supply. Both a dynamic trip, acting as a current limiting circuit, and an overvoltage trip (crowbar) are used to protect the horizontal deflexion and e.h.t. generator. This supply, although complex, offers high efficiency and a very short time-constant, which means that under overload conditions the voltage can be removed from the load in a few microseconds.

The 8500 chassis uses a thyristor (s.c.r.) regulator where the mains frequency is used as the switching frequency (Fig. 5). The choke, in series with the mains, limits the peak value of the repetitive current pulses and attenuates considerably the interference normally generated by the fast switching of the s.c.r. The diode in series with the s.c.r. reduces the peak inverse voltage specification of this device and represents a more economical solution.[4, 5]

In addition to this main supply a 45 V unregulated supply derived from a small mains transformer provides power for the field output stage and a 25 V series regulator supply to the small signal circuits. The 25 V supply acts as a reference for the main high voltage supply.

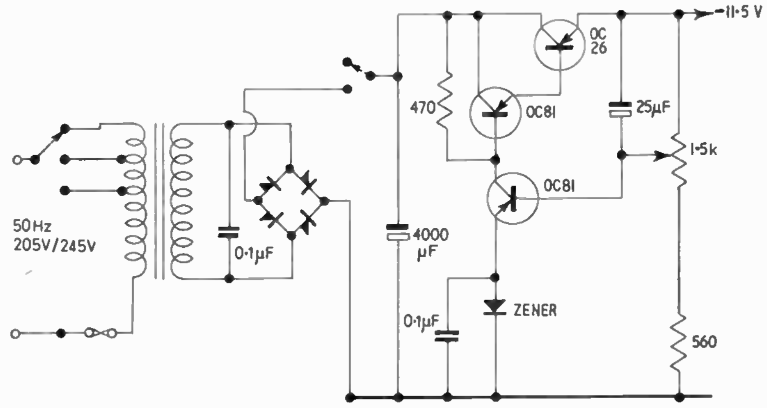


Fig. 3. Power supply for Perdio *Portarama*.

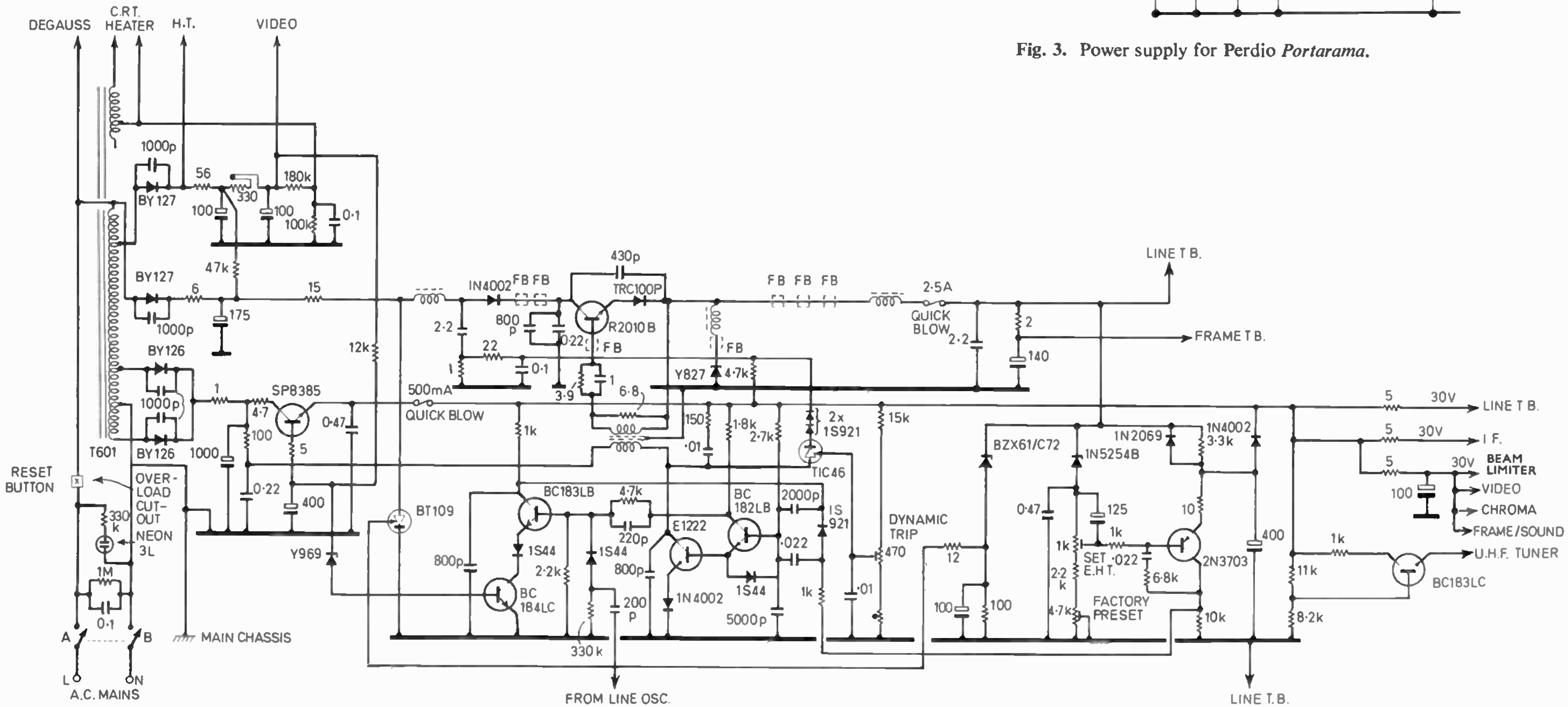


Fig. 4. Power supply of 3500 series.

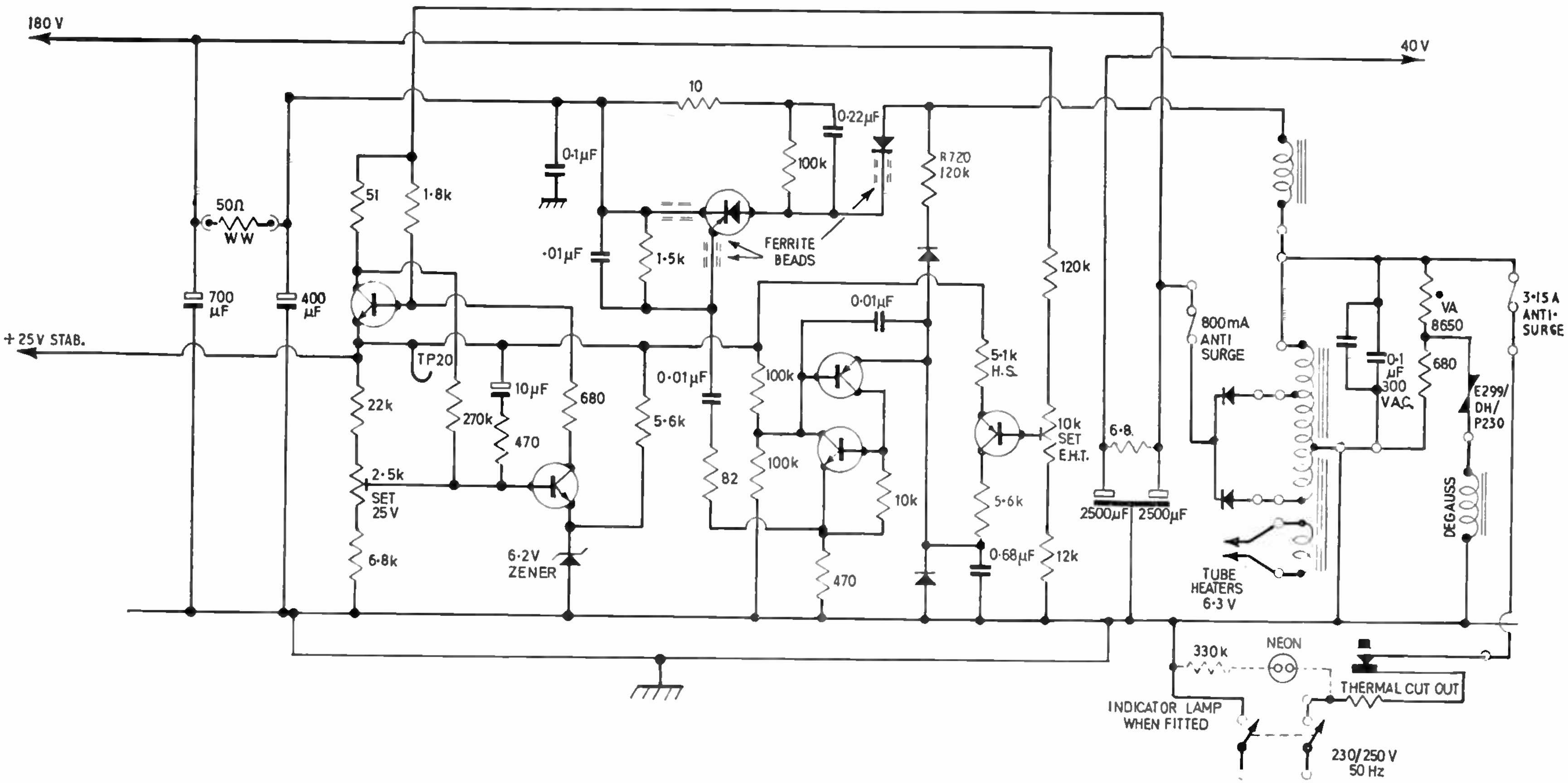


Fig. 5. Power supply of 8500 series.

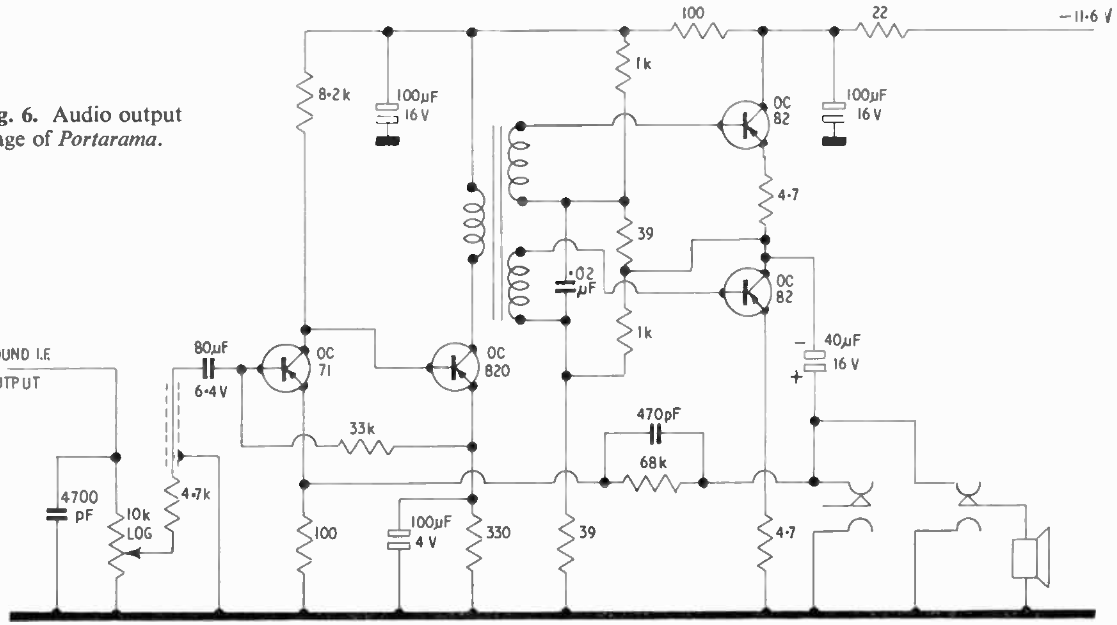


Fig. 6. Audio output stage of *Portarama*.

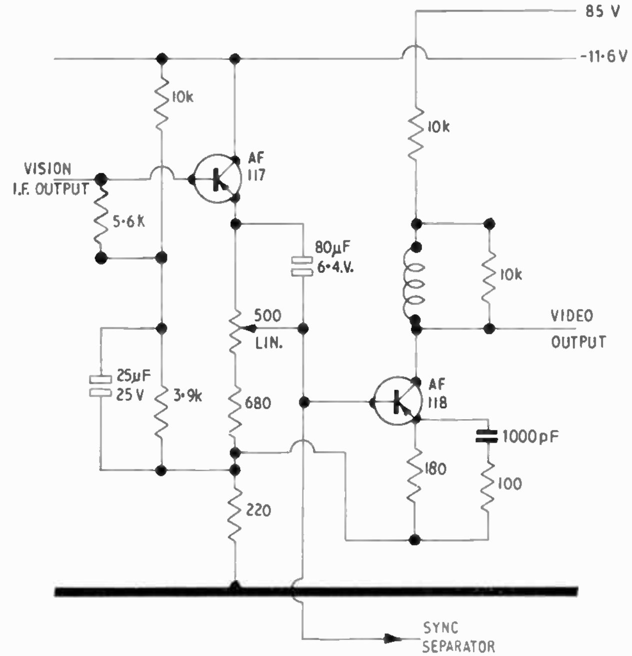


Fig. 8. Video output stage of *Portarama*.

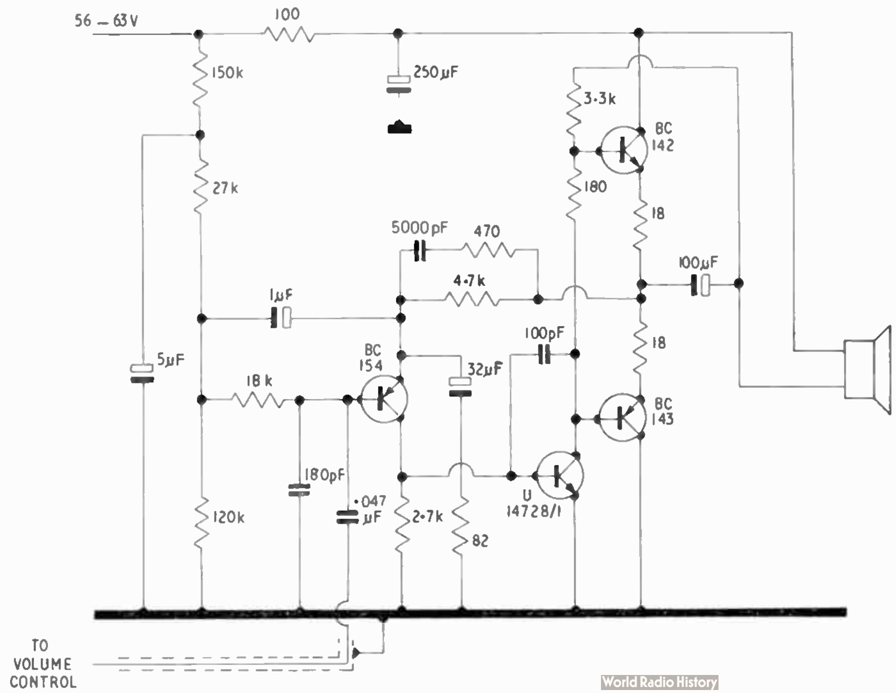


Fig. 7. Audio output stage of 3500 series.

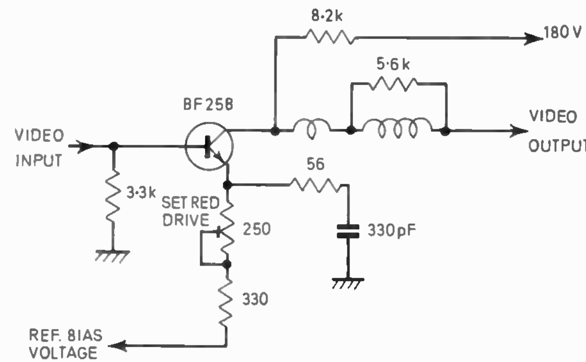


Fig. 9. Video output stage of 8500 series.

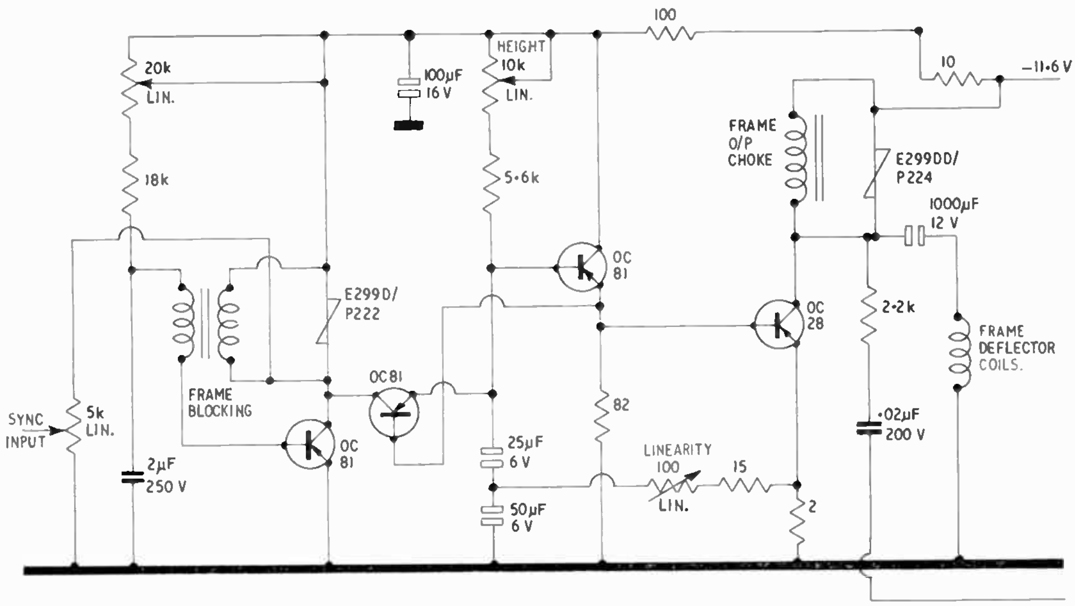


Fig. 10. Field output stage of *Portarama*.

5 Audio Stages

Figure 6 shows the *Portarama* audio output stage. The transformer employed to drive the two output transistors represented an economical solution before the technique of the complementary n-p-n/p-n-p output stage became available to the circuit engineer. Figure 7 shows the audio output stage of the 3500 series with the more conventional complementary stage and without the need for any wound components.

6 The Video Output Stage

An exception to the general trend in circuit development shows that the video output stages have changed little, as Figs 8 and 9 indicate. In this instance improvements in transistor specification have enabled the same basic circuit to generate the required performance in the early *Portarama* and the current 8500 series. Table 1 gives some idea of the parameters achieved in both circuits.

Table 1

Parameter description	*Portarama*	8500 series
H.T. rail	85 V	190 V
Video swing	65 V	140 V
Circuit bandwidth	2·5 MHz	5 MHz

7 Field Timebase

Another example of circuit design with the 'minimum number of transistors' approach is shown in Fig. 10. In this circuit of the early *Portarama* a single transistor blocking oscillator is employed which acts as a clamp during the field retrace period across a simple sawtooth generator. The output stage is driven by an emitter follower and a choke-capacitance coupling configuration is used to drive the field scan coils.

The use of the choke avoided the need for another transistor and typified the stage of the art before the complementary field output stages were developed. It is worth noting that due to limitations in the high voltage characteristics of the transistors available, voltage dependent resistors were used in conjunction with both the blocking oscillator and the output stages.

By contrast Fig. 11 shows the 8500 series field oscillator and output stages. In this circuit wound components have been completely eliminated, silicon transistors have replaced germanium types and time constants are more accurately defined by means of R-C networks.

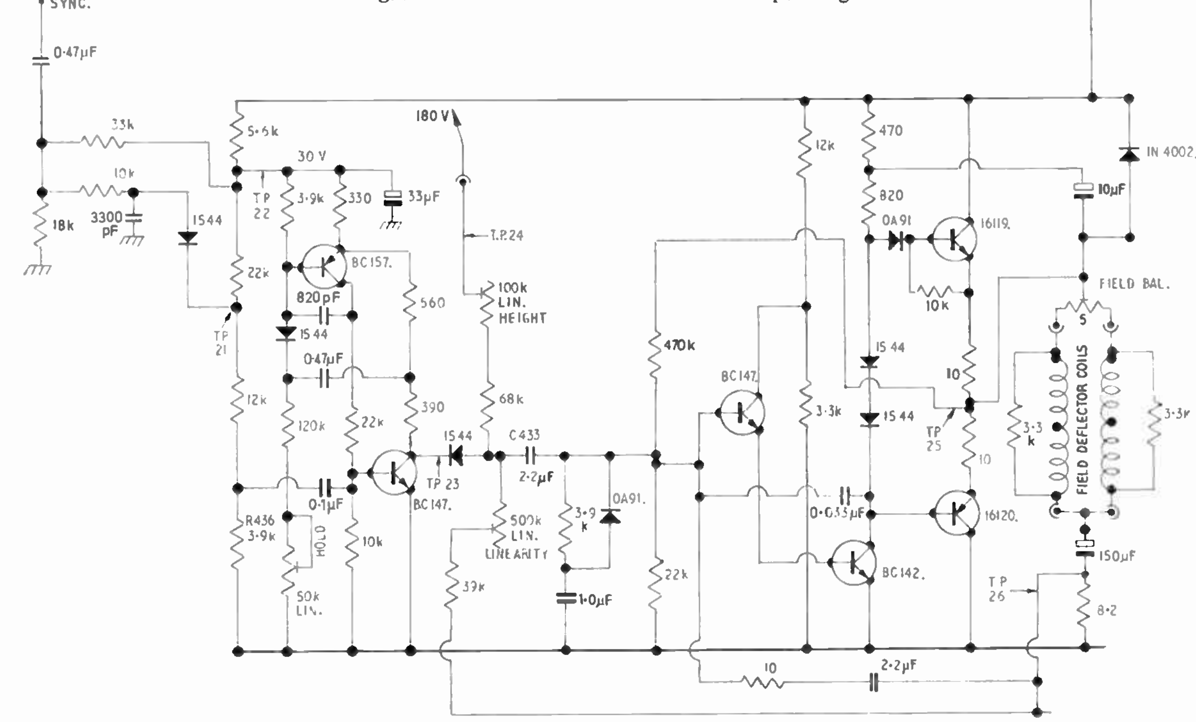
The regenerative action of the two-transistor configurations employed in the oscillator circuit was completely unknown in early designs, where the transistor circuits represented very much a direct translation of the equivalent valve circuits. The number of diodes and transistors employed in the 8500 series timebase circuit would have been unacceptable on cost basis in the *Portarama*.

Similarly the frequency stability versus temperature of the 8500 circuit (0·5 Hz/30 degC) would have represented an impossible target in early designs.

8 Line Timebase

Figure 13 shows the line timebase of the 8500 chassis up to, but not including, the line output stage. In this circuit a transistor has been employed both to invert the sync and to provide some isolation. The rest of the circuit has the same basic configuration as that employed in the *Portarama* (Fig. 12).[6] In both instances an oscillator is frequency controlled by a reactance transistor and the output of the oscillator used to feed the line output transistor driver. In the *Portarama* a blocking oscillator is employed using very simple circuitry although the construction and complexity of the blocking oscillator transformer would not be considered a good production proposition to-day. By contrast the circuit of the LC oscillator in the 8500, designed for a high order of frequency stability, contains many more components but a very simple oscillator winding. In this design no line-hold adjustment is provided as a customer control, as a frequency stability of 1 Hz/degC has been achieved.

Fig. 11. 8500 series field oscillator and output stages.



9 The Line Output Stage

The most dramatic improvement in transistor specifications can be seen by comparing the line output stages shown in Figs. 14 and 15. In the latter the voltage capability of the line output transistor has now reached 1750 V with the BDX32 type of transistor.

In the case of the 8500 series the basic line output transistor stage of the *Portarama* has been simplified further by the omission of the efficiency diode, as the new generation of line output transistors can perform this function by reverse conduction during the scan period.

With the BDX32 series of transistors the switch-off characteristics of the drive waveform are very carefully controlled in order to optimize the delay and storage time of the base emitter diode at the start of the flyback period. The complex network employed in the 8500 helps to achieve high switching efficiency which results in very low running temperatures.

In the early *Portarama* a thermionic e.h.t. rectifier had to be employed purely on economic grounds, although both silicon and selenium diodes were available at that time capable of meeting the required voltage

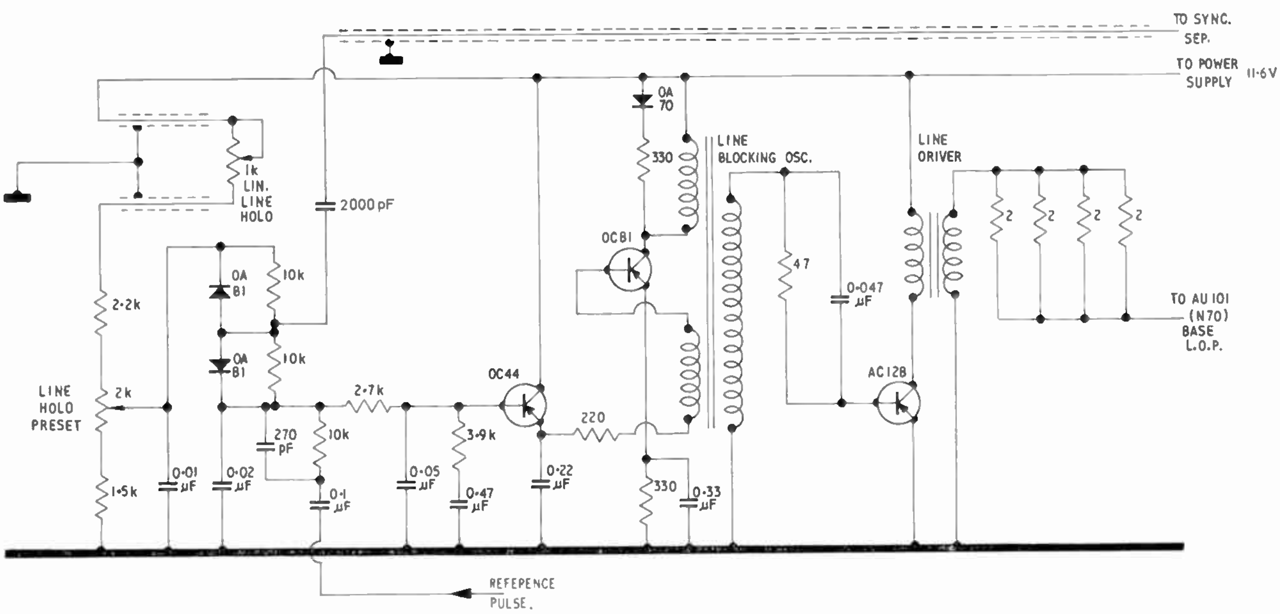


Fig. 12. Line oscillator and driver of *Portarama*.

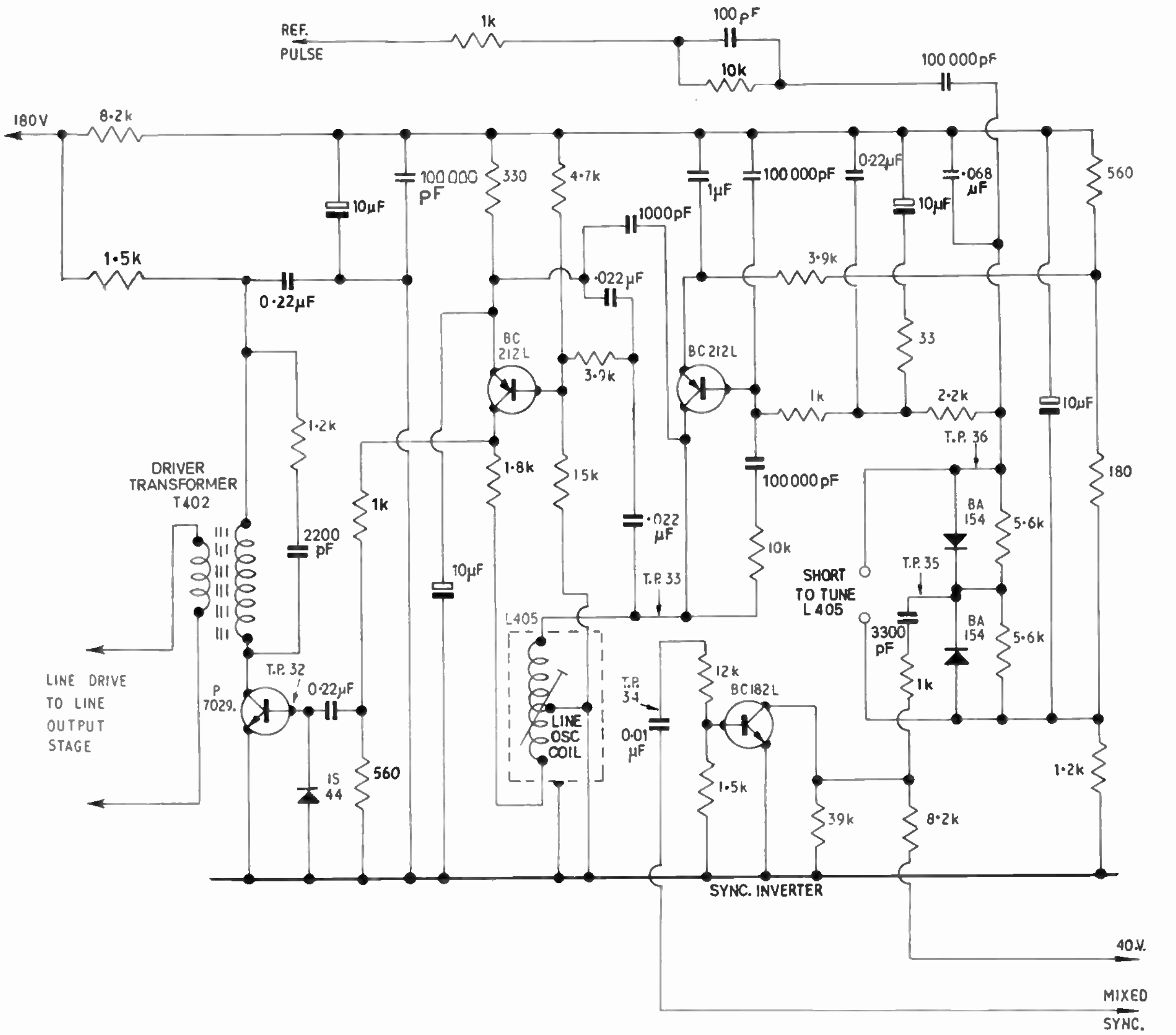


Fig. 13. Line oscillator and driver of 8500 series.

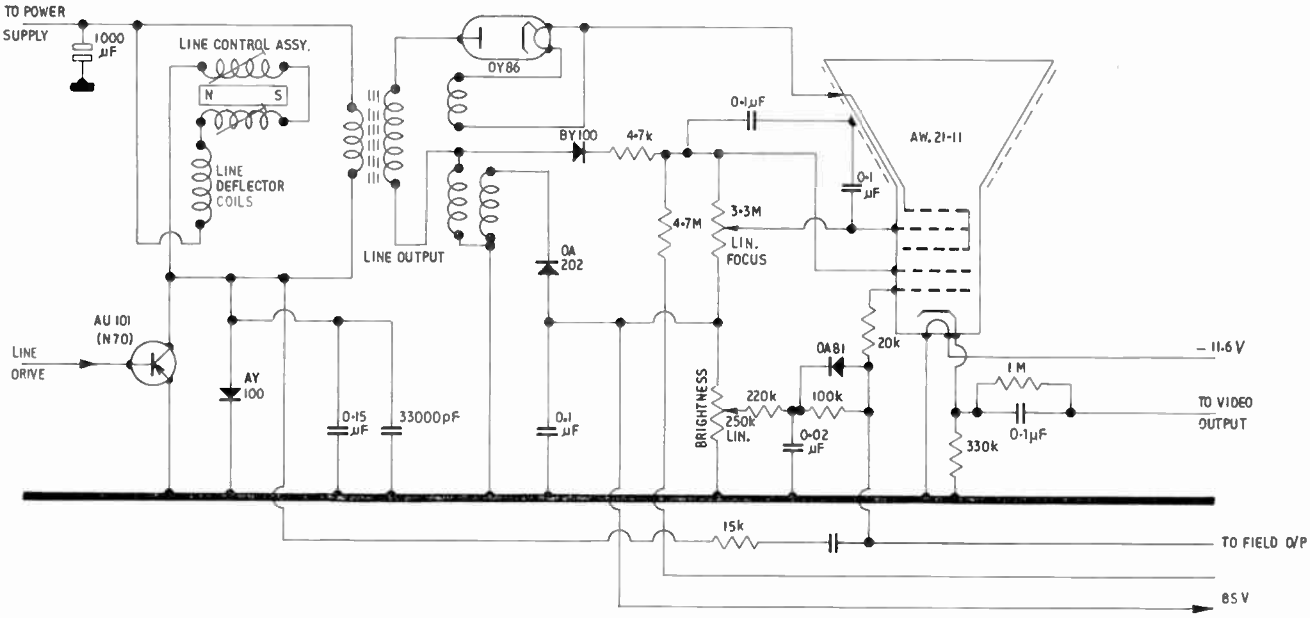


Fig. 14. Line output stage of *Portarama.*

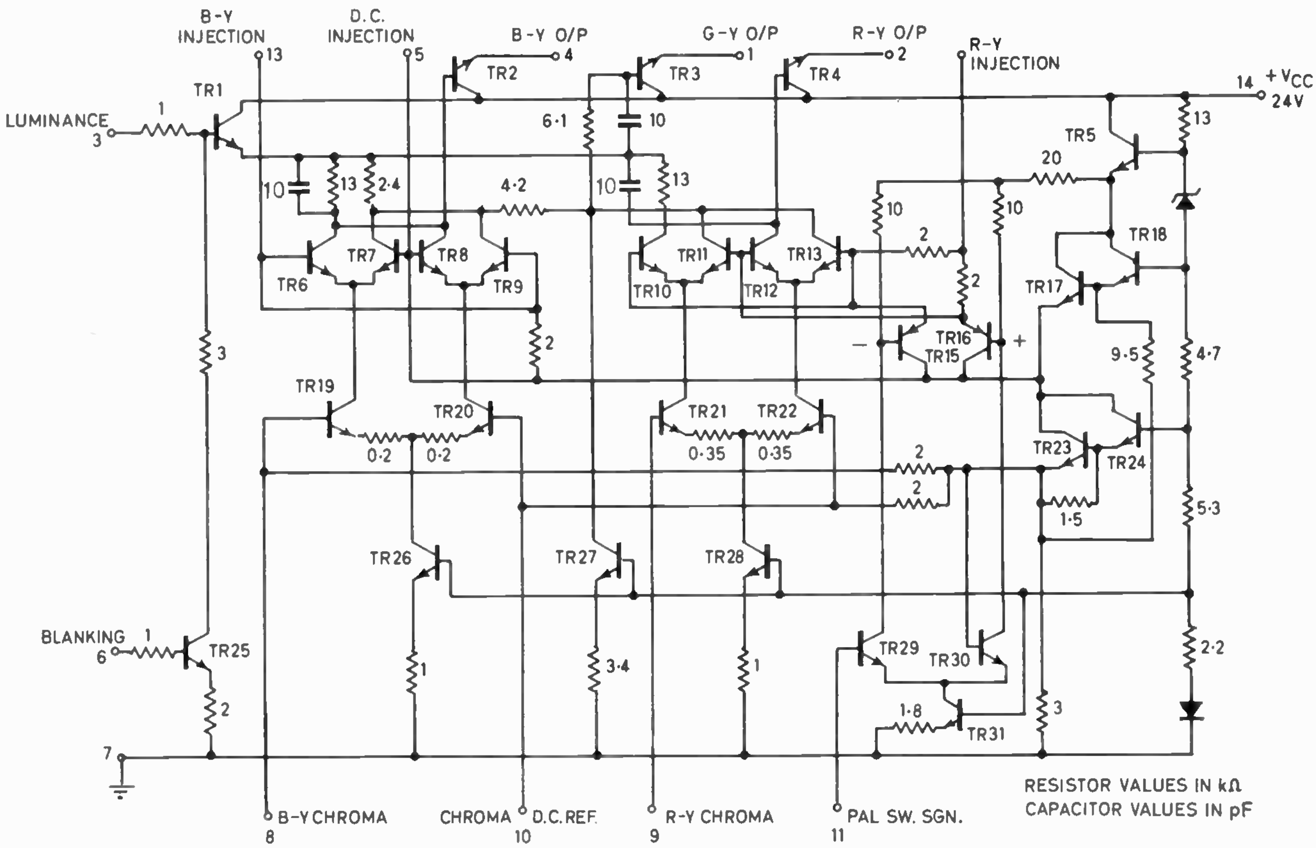


Fig. 18. Chroma demodulator i.c. for 8500 series. (Motorola MC1327P, Texas Instruments SN76227N/07.)

(2) A high order of linearity avoiding the need for a second detector for the chrominance and intercarrier sound components. This linearity offers very low sound/chroma intermodulation products and avoids the need for deep sound trapping in the i.f. filter response.

Figure 16 shows the circuit diagram of this integrated circuit. A video pre-amplifier follows the detector and provides d.c. coupled video signals of 3 to 4 V peak to peak amplitude, plus chrominance signals and intercarrier sound. The d.c. component of this signal can be fully utilized for the derivation of a.g.c. An amplitude-limited i.f. carrier component is used for the generation of a.f.c.

The 8500 chassis uses two more integrated circuits (Figs. 17 and 18). An intercarrier sound i.c. contains limiting amplifiers, quadrature coincidence detector, variable gain amplifiers as well as audio preamplifier. The high level audio output available from this i.c. can drive a single stage class A audio output amplifier.

The third integrated circuit accepts chrominance signals, subcarrier reference signals, half-line indentification signals, luminance and blanking pulses, which after processing provide fully matrixed *RGB* signals to drive the video output stages. Fully-balanced demodulators are incorporated which provide the $R-Y$ and $B-Y$ signals, these being matrixed to obtain $G-Y$ inside the i.c. The luminance (Y) signal is then added and the resultant *RGB* signals formed at three low impedance outputs to drive the video transistors direct. Since the demodulators are balanced there is very little subcarrier present at the outputs and the higher-order demodulation products (predominantly 2nd harmonic) are reduced by RC filtering on the i.c. chip.

Television receivers are now available with integrated circuits fulfilling many more functions than those described above. This is perhaps the biggest example of the progress which one decade of semiconductor technology has brought to the television receiver: a technology which started much earlier with a crystal detector.

It is difficult to conceive, judging by this rate of progress, what another decade of semiconductor technology will mean to television and what improvements in performance this will represent. One thing is certain: it looks exciting indeed!

11 References

1. Perdio *Portarama* service manual.
2. B.R.C. 3500 colour television service information.
3. B.R.C. 8500 colour television service information.
4. 'The Controlled Rectifier,' Vol. 1, I.R. Corporation, El Segundo, California.
5. King, D., 'Stabilized power supply using a thyristor', Mullard Technical Note, 1967.
6. Martin, K. E. and Hughes, J. B., 'Line time bases', paper J—Mullard Technical Publications—May 1961.

Manuscript first received by the Institution on 16th October 1972 and in revised form on 6th November 1972. (Paper No. 1502/Com 62.)

UDC 621.31

The impact of semiconductor devices on electrical power engineering

Professor J. J. BATES,
B.Sc., Ph.D., C.Eng., F.I.E.E.*

and

R. E. COLYER,
B.Sc., Ph.D., C.Eng., M.I.E.E.*

SUMMARY

The differences in approach between power engineering and electronic engineering are briefly outlined and the role of semiconductor devices in place of those based on the mercury-arc is discussed. Applications of the diode and thyristor to d.c. and a.c. machines are described together with the use of thyristors in high voltage d.c. power transmission. The importance of careful integration of semiconductor devices with the associated machines in any proposed scheme is emphasised.

* *Electrical Engineering Branch, Royal Military College of Science, Shrivenham, Swindon, Wilts.*

1 Some Important Differences Between Power Engineering and Electronic Engineering

In this busy day and age, many of us engaged in technological work seldom have time to pause and think about the way particular innovations shape the subject speciality loosely called our 'profession'. But it sometimes pays to pause and reflect, because by realizing the circumstances that have brought new devices into use one is better able to appreciate the way things may progress in the future.

It is often quite remarkable today how people who have played no part in the development of a new device, and who often have little appreciation of how it *really* works, can utilize it to great advantage. A very good illustration of this is in the impact of semiconductor devices on power engineering. The power engineer, fabricating devices from mainly copper and iron, using largely mechanical skills, is not the man easily to appreciate the solid-state physics on which semiconductor devices depend. Similarly, the semiconductor engineer has insufficient knowledge of the problems of economically successful power engineering (though he may more easily grasp the simpler principles of action of machines) to apply semiconductor devices in such a field. It is thus in some ways remarkable that power semiconductor devices have been successfully applied to electrical power and machines, and understandable that this has taken a considerable time.

Power engineering is basically the technology of energy conversion and utilization; essentially the electrical extension of the mechanical power sources that originated with the industrial revolution. The attitudes and approaches demanded from a power engineer are very different from those of an electronics engineer, because the former is so very concerned with large amounts of energy which he must always try to conserve, never lose. Energy is not a prime consideration in electronics, and most of the ingenious signal manipulations are accomplished because the energy levels are sufficiently low to allow the incorporation of energy dissipating devices (i.e. resistors) to a great extent. On the other hand resistance in power engineering is usually looked upon as nothing but a nuisance, to be avoided at all costs. There is a very simple explanation for this: in any 'lossy' electrical device, the total energy loss will vary as the cube of the dimensions, i.e. as the volume, whereas the area available for heat dissipation varies as the square of these dimensions. Hence as size increases the ability to dissipate heat is rapidly outpaced by the rate at which heat is generated and in dealing with large electrical devices efficiency is a prime consideration. Energy must always flow from one part of a system to another with the absolute minimum of loss, and a machine system just would not be possible if the efficiency was not high; it is not just a question of improved performance. Another important difference between power engineering and electronic engineering is the sheer damage that can be caused by circuit faults. The electronic circuit, with its multiple incorporation of high impedance paths means that fault power is largely inherently limited when a component fails. But the power circuit, with very low impedance paths cannot have this inherent in-built protection. These are very important factors when trying to persuade engineers to use what seem to them delicate semiconductor devices relative to the copper and iron with which they usually work, and the failure of which could lead to destructive currents. This often implies that the protective circuits that have to be incorporated may be extensive and expensive, and reliability and simplicity are of prime importance.

2 The Evolution of Mercury-Arc Devices

From the earliest days of electrical technology it has been appreciated that electricity was a 'force' that could be easily controlled compared with mechanical forces. Put in a crude way a switch is much simpler than a mechanical clutch. But what electrical engineers did not have to quite the same extent was a means of varying their powers progressively in any simple way especially where d.c. was concerned. They could easily have it or not have it, switch it on or switch it off, but nothing in between. The mechanical engineer could control the power of his steam engine progressively by simply using a valve to control the amount of his working fluid passing. The electrical engineer's simple equivalent was often limited to a variable resistance, but this only controlled by means of large power losses and the invention of the thermionic valve, though controlling electricity with electricity, was really no different. Though the power engineer must have watched with envy the fascinating ways in which this new device was put to use, he knew that he himself could not use it; he could see straight away that it was essentially an electrically-controlled resistance—a dissipating device—and dissipating devices were out as far as he was concerned. His controlling devices had to be open with no current and no loss, or closed with no voltage drop and no loss.

The first controlled switch to have any real impact on power engineering came with the development of mercury-arc devices, and particularly the grid-controlled mercury-arc rectifier. This was a controlled valve, capable of handling high powers. It was still not a valve with the versatility of the electronic triode, because it was a device that had to be used fully on or fully off. But though there was no progressive control there was the facility to switch on and off at repetition rates that could not be matched by any mechanical switch.

Thus the power engineer came to grips with a device whereby he could control electrical energy by letting it through in bursts. Because many of the applications for which electrical power is ultimately needed involve mechanical drives with relatively high inertia, the fact that the energy came in a pedal/freewheel manner was of little importance. The minds of (some) power engineers thus turned to rapid switching as a means of control and static voltage control systems and d.c. to a.c. inversion became technically and economically feasible in some cases using mercury-arc devices.

Mercury-arc devices however were rather bulky and the high voltage drop in the arc made them inefficient on lowish voltages. Their economic use was therefore restricted to bulk a.c. to d.c. conversion, the armature voltage control of large d.c. machines and the special field of electric traction.

3 The Semiconductor Equivalents of Mercury-Arc Devices

The situation changed dramatically with the development of power semiconductor devices and the arrival of highly reliable germanium and silicon diodes of high power rating in the '50–'60 decade made the local small-scale conversion of a.c. to d.c. quite feasible and economic. But it was the quickly following thyristor that really marked the threshold for the introduction of semiconductor devices into power engineering, because it made possible the replacement of the generator in the familiar Ward-Leonard d.c. motor control systems by a relatively simple, cheap and very compact static device. In other words it provided a variable voltage d.c. power supply that was capable of accepting regenerative power and, most important of all, could do this at ratings which encompassed many of the drives that currently employed Ward-Leonard systems.

The impact of the thyristor was thus principally because there was an easily-seen large-scale application which was not overcomplicated and the ultimate success of which was fairly easily predictable. Its impact has been much greater than that of controlled mercury-arc devices for a number of reasons. It is a much more efficient device, quite suitable for low-voltage applications, and is simpler, smaller and very reliable. But the impact has been greater for subtler reasons than this. The thyristor was from the early days a readily obtainable device—a stock item. It has been played with in industrial laboratories and universities to an extent that was never the case with mercury-arc devices. In academic institutions, where electrical power work has steadily lost ground to the more appealing and attractive fields of electronics, it was seen as a means of redressing the balance and bridging the gap between electronics and power engineering. The importance of such interest-creating factors should not be casually dismissed; their effect on the rapid application of the thyristor has been far reaching.

Replacement of d.c. generators in Ward-Leonard systems is not of course the only application of the thyristor. Many other ingenious uses have been found for it, ranging from d.c. to a.c. inversion to d.c. machines without commutators. Some of these are economically and technically viable now, and more may become so in the future. The following sections of this survey paper deal with particular applications.

4 Diode Applications

The first semiconductor device to be used in power engineering was the high power diode which became commercially available in the early '50s, initially based on germanium but very soon on silicon. This device was a very convenient replacement for arc devices such as ignitrons which were at that time being used in commercial rectifying equipment and its smaller size and fewer auxiliaries allowed considerable savings to be made in equipment size and weight. This was particularly important for portable equipment such as that used on a.c. locomotives.[1] Large banks of diodes are now also used in traction schemes for bulk d.c. supplies, as a replacement for multi-anode steel tank mercury-arc rectifiers, a typical illustration being the many 3000 kW, 3000 V substations of the Belgian State Railways.

The small size and robustness of the semiconductor diode allowed these devices to be built onto the rotating parts of electrical machines, leading to the brushless alternator.[2] Here the rotor excitation is provided by rectifying, on the rotor, the alternating voltages obtained from an integrally rotating a.c. exciter with a stationary field winding. This application is now widespread.[3]

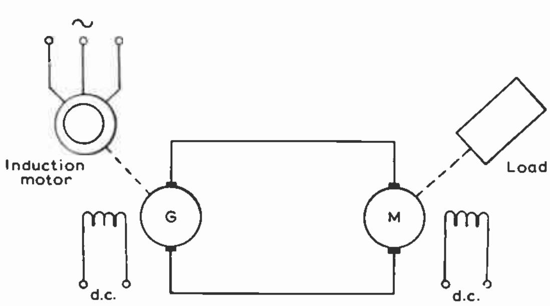
A major requirement for low voltage d.c. has always existed in the electroplating industries, and in the electrolytic refining of aluminium. Neither of these lends itself economically to mercury-arc rectification because of the high arc drop. Shortly before the advent of the semiconductor diode considerable development of mechanical type rectifiers took place, where mechanically-driven switches are operated in synchronism with the a.c. supply. This development was virtually wiped out overnight by the semiconductor diode, where the low voltage drop made efficient high current rectification by extensive paralleling of diodes feasible.

5 Thyristor D.C. Drives

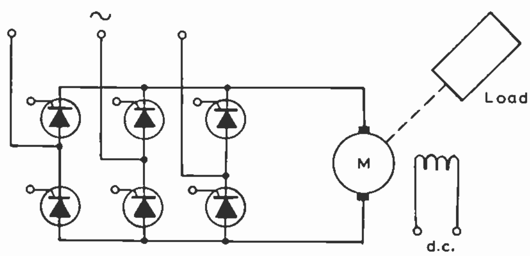
5.1 *Ward-Leonard Equivalents*

The replacement of d.c. generators in Ward-Leonard systems by thyristor convertors has already been mentioned as one of the most important applications of semiconductor devices to power engineering. This was the first application of thyristors to gain general industrial acceptance and probably accounts for the majority of the installed capacity of thyristor equipment at the present time. There are several reasons for this, one of the principal ones being that the change from a rotating to a static power amplifier involved little change *in principle* and allowed the same familiar and well proven d.c. drive motor to be used. Furthermore, the conversion of a.c. to variable d.c. as performed by the phase-controlled thyristor in this application involved natural commutation[4] by the a.c. supply and hence avoided sophisticated auxiliary commutation circuits and required thyristors of only modest capabilities. There are, however, considerable advantages in terms of improved performance, reliability and reduced installation costs to be gained from using the static system.

Figure 1 shows in outline form the general arrangement of a traditional Ward-Leonard system and its modern thyristor counterpart where the motor-generator set is replaced by a thyristor convertor. This set contains two machines of equal size and rating to the drive motor, adds considerably to the total equipment weight and hence to the civil engineering costs of a large installation, and introduces triple energy conversion into the system. Speed and torque control are obtained by varying the current through the field winding of the main generator and since this winding is always highly inductive, a limit is automatically set on the fastest response which can be obtained with such a system. With the thyristor convertor, however, the civil engineering requirements are modest, there is no triple energy conversion and no such limit on response because the mean value of the d.c. voltage from the convertor is controlled by varying the firing point of the thyristors in each half cycle by purely electronic means. Moreover, since the power required to control the firing circuits of the convertor is so much lower than that required to control the excitation of the Ward-Leonard generator it is possible to apply all the modern techniques of signal processing to the system and so satisfy the very demanding performance specifications demanded by many present-day applications.



(a) Ward-Leonard system



(b) Thyristor counterpart of Ward-Leonard system.

Fig. 1. Variable speed d.c. motor drives

But it would be unfair to pretend that there are no problems encountered with such thyristor drives. There is no substitute in the thyristor drive for the stored rotational energy of the Ward-Leonard system which gives the latter an ability to accommodate sudden changes in the drive motor load without drawing an excessive current from the supply. The large difference in thermal time-constants between a machine and a thyristor also demands extensive protection circuits in the thyristor convertor. There are also problems associated with the unidirectional conducting property of a thyristor convertor which, although it can accept regenerative power by operating in an inverting mode,[5] can drive current in only one direction through the drive motor and hence produce rotation in only one direction. If a reversible thyristor drive is required it is therefore necessary to reverse either the field or the armature connexions of the drive motor or, if this is unacceptable for reasons of response time, to go to an arrangement with two inversely-connected convertors.[6]

5.2 *Chopper Regulators*

The thyristor equivalent of the Ward-Leonard system is essentially one of controlling the mean value of a rectified a.c. supply by varying the firing point in the sinusoidal waveform.

A chopper regulator can also provide a variable d.c. voltage but starting from a fixed d.c. voltage. It consists basically of a switch connected between the supply and the load which is opened and closed at high frequency and where the variation of the open to closed time of the switch (the mark/space ratio) varies the mean value of the voltage supplied to the load.

The switching operation can be performed very effectively by a thyristor but since the thyristor is connected into a d.c. supply it must be turned off at the appropriate times by a forced-commutation arrangement. This operates by forcing the current through the thyristor to zero when turn-off is required and then holding it reverse biased for a sufficient time for it to recover its reverse blocking characteristics. There is a very large number of possible chopper regulator circuits[7] differing mainly in the commutation arrangement but most use the energy stored in a capacitor either directly or in conjunction with inductors to perform this task. Since there may be a large number of commutations per second, the capacitor and the whole commutation circuit is extremely important and may well be the main determining factor in the success or otherwise of the complete chopper circuit. The basic form of a typical chopper regulator circuit which has found considerable application in practice is shown in Fig. 2.

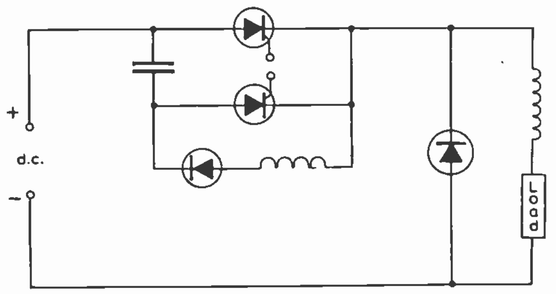


Fig. 2. A basic chopper regulator circuit.

It is important to note that if the load contains inductance, chopping the d.c. voltage does not mean that the current supplied to the load is simultaneously chopped since the essential free-wheeling diode connected across the load allows current to continue to flow through the load after the supply has been chopped. Under these conditions, if losses are neglected, the ratio of the mean current from the supply to that through the load is equal to the ratio of the mean load voltage to supply voltage and the chopper behaves exactly as a variable ratio d.c. auto-transformer.

Chopper regulators have general use in many types of d.c. power supply but have been found to be particularly suitable for the control of the motors in battery electric vehicles where their increased efficiency over that of series resistor controllers allows increased range to be obtained from a given battery capacity. It is possible to obtain an increase in voltage level using a slightly different arrangement of chopper circuit[8] and this principle is used to obtain regenerative braking in chopper-controlled battery vehicles as a means of obtaining still greater range from the battery. Large d.c. machines will in any case invariably obtain their power from a.c. supplies and the simplicity of natural commutation relative to forced commutation means that chopper regulators are restricted to the control of small to medium-sized machines.

6 Thyristor A.C. Drives

6.1 *D.C. to A.C. Inversion at Variable Frequency*

The presence of the commutator in a d.c. machine makes it expensive, gives it a poor power/weight ratio and imposes severe limitations on the maximum power and speed at which it can operate. But the d.c. machine possesses great inherent controllability and is able to operate at high efficiency over a wide range of conditions.

On the other hand, a.c. machines, particularly the squirrel-cage induction motor, are simple and cheap and suffer none of the limitations associated with a commutator but they have the very major drawback that their speed is closely related to the supply frequency which until recently was not easily variable. There has always been considerable incentive to produce variable speed a.c. machines, however, which could operate directly from the fixed-frequency supply system and considerable ingenuity has been employed to this end. But the resulting machine systems are usually very complex and also require expensive commutators and brush-gear[9] with greater associated problems than the equivalent d.c. machine. The reason for this situation is simply that any variable speed a.c. machine must contain a frequency-changer of some form or other and prior to the arrival of the thyristor the only form of frequency changer was a rotating machine.

The thyristor has greatly changed this situation however, since with the introduction of fast turn-off thyristors and the development of special capacitors for use in commutation circuits, it is now possible to design efficient invertors which can convert d.c. to a.c. of fixed or variable frequency. The d.c. may be obtained from a rectified a.c. supply of fixed frequency and the complete system then becomes a frequency changer. Since the thyristors in such an invertor operate from a d.c. supply like those in chopper regulators and there may be no live supply on the a.c. side to perform natural commutation they must again be turned-off using a forced-commutation arrangement. There are many possible invertor circuits[10, 11] which again differ mainly in the commutation arrangement and Fig. 3 shows in outline one three-phase invertor circuit which has found considerable application in practice, the diodes being included so that the invertor can supply inductive loads.

The basic action of most invertors produces phase voltages having a square waveform and the combination of these with appropriate displacements in a three-phase arrangement yields one of the now characteristic line-voltage waveforms shown in Fig. 4. A.c. machines both synchronous and asynchronous operate quite satisfactorily from supplies having these voltage waveforms above a frequency of a few hertz with a degradation in performance of less than 5% under reasonable

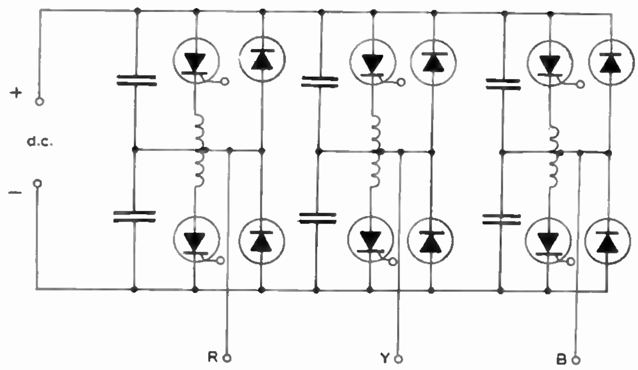


Fig. 3. A basic circuit of a forced-commutated three-phase inverter

conditions; thus, since it is possible easily to adjust the frequency of the invertor, the power engineer's dream of an efficient variable speed a.c. machine system has now become feasible. It is normally necessary to hold the air-gap flux-density in machines at a constant value at all operating frequencies and this is achieved by varying the voltage in proportion with the frequency, thus maintaining a constant voltage/frequency ratio. Voltage changes may be obtained either by variation of the d.c. level on the supply side of the invertor using a thyristor convertor to supply the d.c., or by interposing a d.c. chopper between a fixed d.c. supply and the invertor and both systems are used in practice.[12] Alternatively in some types of invertor, voltage variation may be achieved by varying the width of each half-cycle of the output voltage.

A technique which has become possible with the development of very short turn-off-time thyristors is that of pulse width modulation (p.w.m.) in which the voltage waveform is synthesized from a number of pulses of varying width as shown in Fig. 5 in such a way that the resulting waveform has a very low harmonic content. Since it is possible to vary the width of all pulses within each half cycle while maintaining the relationship between the width of successive pulses it is possible to obtain both voltage and frequency variation within the invertor which can then operate from a d.c. supply of fixed voltage. This arrangement greatly simplifies the design of the d.c. supply and allows the connexion of an auxiliary supply

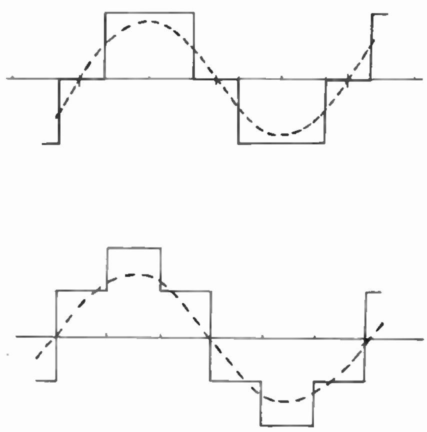


Fig. 4. Typical voltage waveforms of three-phase inverters.

for 'no break' facilities. The low harmonic content of the voltage waveforms obtainable using a p.w.m. type of invertor allows operation of machines at very low speeds and such invertors are finding increasing use in reversible a.c. drives.[13]

If an invertor is used to supply a synchronous machine then the accuracy of speed is determined purely by the maintained accuracy of the oscillator driving the invertor, which can be of a very high order. An arrangement consisting of an invertor supplying a number of synchronous machines is particularly useful where exact synchronism of a number of drives is required while the speed of the whole system is varied and such a specification would be difficult to satisfy in any other way. Transfer rollers in steel mills, and certain textile drives use such multiple drives.

When induction motors are supplied from an invertor the speed of operation depends on the slip frequency as well as the frequency of supply. Using a system in which the slip frequency is controlled, however, it is possible to obtain full-load torque from the machine over a very wide range of speed and the resulting characteristics are very similar to those of a d.c. machine supplied from a

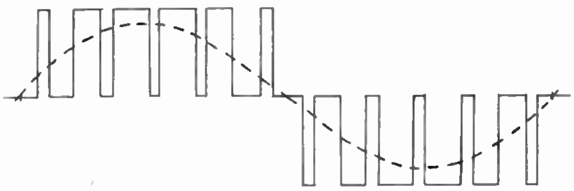


Fig. 5. Voltage waveforms of p.w.m. inverter.

Ward-Leonard system or a thyristor convertor. This similarity may be extended to include regenerative braking if the induction motor is allowed to operate with negative slip provided that the d.c. supply is able to absorb this power. It may be seen then that in situations where either the speed of operation is higher than could be attained with a d.c. machine or where a commutator cannot be tolerated it is now possible to use an a.c. machine supplied from an invertor and this arrangement also opens up the possibility of obtaining very high power/weight ratios from machines by operating them at high speed. Already speeds in excess of 50,000 rev/min have been achieved in some special applications.[14]

D.c. to a.c. inversion is not of course only required for driving rotating machines of the induction or synchronous type. The operation of electronic equipment from d.c. supplies including batteries, such as for instance must be the case in military vehicles on silent watch, may require powers of upwards of 1 kW to supply communication equipment and computers. If a distinction does exist in this use of inversion compared with the supply to machines, it lies in the trouble taken to filter the basically square-wave output to obtain a good sine-wave supply of limited harmonic content. Certainly the filtering equipment in such invertors is often larger than the invertor itself. This is in some ways an odd situation, because at some stage in the overall system the a.c. is invariably rectified back to d.c. at an appropriate level,

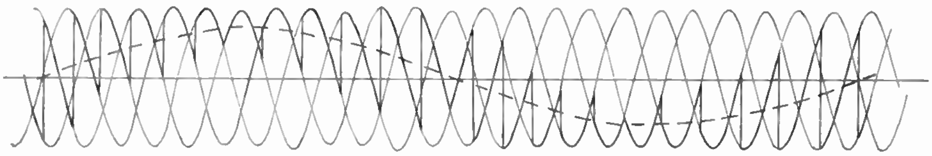


Fig. 6. Synthesis of a l.f. waveform from a h.f. waveform in a cycloconvertor.

and yet good ripple-free d.c. is most easily obtained from rectified square-waves rather than sine-waves. One can reflect that all the bulky filtering equipment built into such invertors to give a good a.c. waveform is effectively repeated elsewhere, albeit in small packets, to reverse the process!

6.2 *Cycloconversion*

The principle of cycloconversion to provide variable frequency a.c. from fixed frequency a.c. was well established in the early '30s using mercury-arc rectifiers[15] and has been used extensively in continental railway systems as a means of converting the 50 Hz national supply to a supply at 16$\frac{2}{3}$ Hz for single phase a.c. traction motors. The basic principle consists of synthesizing a low frequency voltage waveform from sections of a high frequency waveform, as shown in Fig. 6. In practice this is achieved using two inversely-connected convertors which are allowed to conduct alternately and such an arrangement is shown diagrammatically in Fig. 7 where a low-frequency single-phase supply is obtained from the higher frequency three-phase supply. Consideration of Fig. 7 indicates that for a system to give a three-phase output the number of rectifying devices will be very high and this is one of the disadvantages of the cycloconvertor which until the advent of the thyristor made this system quite uneconomic for other than very large static installations. The other major disadvantage of cycloconversion which may be appreciated from inspection of the waveforms of Fig. 6 is that when the converted frequency becomes greater than about one half of the starting frequency the harmonic content of the voltage waveform becomes excessive.

The thyristor has made the system economically viable in smaller sizes and where limited variable speed operation of a.c. motors is required it is finding increasing application,[16] since the system does possess a number of advantages over the invertor systems discussed in Sect. 2.3.1. The cycloconvertor operates under conditions of natural commutation and has an inherent ability to accept regenerative power without any additional auxiliary circuits.

An unusual application of cycloconvertors in which they are finding use in the reverse of their normal operating mode is the generation of constant frequency a.c. supplies from the variable speed alternators in some modern aircraft. The alternators are driven from the main engines and since these are subject to considerable speed variation it was necessary in the past to use a constant-speed drive of some form between each engine and alternator and these drives were heavy and expensive. The recent trend is to drive high frequency alternators directly from the main engines and convert the variable frequency obtained from them into constant frequency by means of cycloconvertors.[17]

7 High-voltage D.C. Power Transmission

There are some situations where it is advantageous to transmit electrical power using d.c. instead of a.c. and a number of such schemes are in existence throughout the world.[18] In the past these schemes have invariably used mercury-arc rectifiers and it has been possible to develop devices capable of withstanding voltages of several hundred kilovolts and of conducting currents of hundreds of amperes. But there appears to be little prospect of producing a single-thyristor equivalent of the high voltage mercury-arc rectifier and in designing a thyristor system for use in h.v.d.c. schemes it is therefore necessary to connect many thyristors in series/parallel chains. When such an arrangement is used, however, problems are encountered in respect of voltage sharing between series-connected thyristors during blocking, and current sharing between parallel-connected thyristors during conduction and these problems have to be solved by adding extensive sharing networks to the assembly. There is also considerable difficulty in supplying firing pulses to the thyristors in the stack which are all at different potential with respect to each other and one successful solution to this problem is to initiate the firing pulses through fibre optics which then allow isolation between the thyristors and the firing circuits.

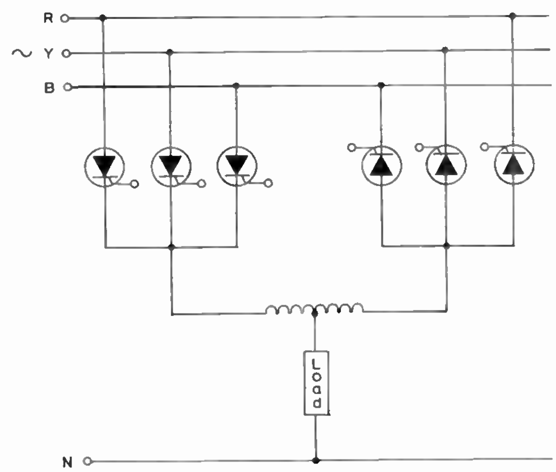


Fig. 7. Three-phase/single-phase cycloconvertor.

These difficulties, together with the excellent record which high voltage mercury-arc rectifiers have achieved as a result of many years of development, have delayed the extensive application of thyristors to a far greater extent than in most other areas of power engineering. But successful, and perhaps what is more important, competitive thyristor valves have now been developed[19] and are being incorporated into h.v.d.c. schemes. A development scheme has been in operation in Sweden for some years and others are under construction in other parts of the world. As a result of the number of thyristors in series in a thyristor valve, the voltage drop when conducting is higher than in an equivalent mercury-arc rectifier, but operating conditions are less stringent and the number of auxiliaries required considerably reduced and these factors together with the greater reliability expected should yield improved overall performance.

8 The Closer Integration of Thyristors and Machines

8.1 *The Effect of Thyristor Supplies on Machine Design*

The applications so far described have been essentially means of obtaining variable voltage and/or frequency, from d.c. or a.c. sources to feed *conventional* d.c or a.c. machines and so facilitate their control and extend their field of application.

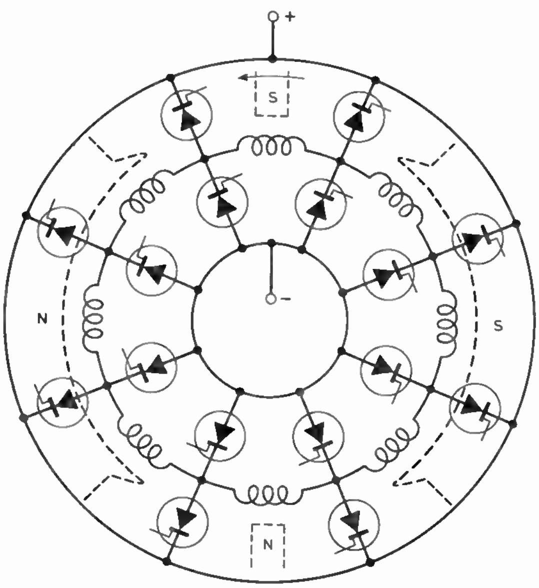


Fig. 8. Commutatorless d.c. machine.

In some cases there is a close interaction with the machine itself that has to be considered. For instance, a thyristor supplies fairly rough d.c. and if this is used to feed a d.c. machine it may be necessary to carefully design the machine itself to allow for this. Thus in the larger machines it may be essential to use fully-laminated field systems, and in all cases it is probably necessary to exercise care as to the design value of reactance voltage and the choice of the most suitable brush grade. The advent of the thyristor has in effect led to a more critical look at even such a well-established article as a d.c. machine and is a good illustration of the way in which the introduction of a new device has repercussions on well-established engineering procedures. Twenty years ago few d.c. machine designers would have given credence to a statement that in the future they would have to look more closely into commutation phenomena because certain developments were taking place in the electronics field.

Similarly, the operation of induction motors from essentially square-wave supplies has led to new ways of calculating induction motor torques, and finding out whether an induction motor fundamentally dislikes, or indeed possibly prefers, a square-wave supply compared to a sine-wave supply. The results of such studies may point the way to general improvements in induction motors that are unrelated directly to the use of such machines from thyristor-controlled sources.

Even such an apparently simple application as the incorporation of rotating diodes in brushless alternators has led to interesting work on over-voltages that can be induced in the stator windings due to sudden loss of excitation, as the failure of a diode on such a machine would correspond to such an abrupt change in excitation that dangerous over-voltages might be created. Damping windings have had to be built into such machines to counter the possibility of abrupt flux changes.

The essential lesson from all this is that the successful application of thyristors to power engineering involves an extensive knowledge of both thyristors and machines. The two aspects cannot be divorced and pursued in relative isolation. There is a need for a new breed of engineer.

8.2 *The Replacement of D.C. Machine Commutators by Thyristors*

The commutator in a d.c. machine performs essentially a switching action as the brushes pass from segment to segment. It is in many ways a remarkable mechanical switch, considering the magnitude of the current switched, the inductive nature of the circuits switched, and the number of switching actions in the life of a commutator.

It is possible to replace the commutator switching action entirely by thyristors and so obtain so-called commutatorless d.c. machines. Each commutator segment must in effect be replaced by a pair of back-to-back thyristors, as each segment has to pass current in both directions during each pole-pair of rotation. A first essential requirement in the thyristor equivalent of a commutator is thus a reduction in the number of armature tappings or commutator segments from a typical value of 50 per pole-pair to some 8 per pole-pair. Even this means that 16 thyristors are required for a two-pole machine, but two-pole machines are not economic in iron and copper utilization, and so the minimum number of thyristors required in any reasonably-sized practical machine must be 32. A second requirement is that the machine must be inverted to have a stationary armature and a rotating field system, so that the thyristors are stationary. This is not as simple as it sounds as interpoles (if used) will be rotating and unless the machine is series-wound, slip-rings have to be provided for both interpole excitation and main-field excitation. The thyristors also have to be triggered from the shaft position.

Such a machine is shown in principle in Fig. 8. The thyristors are fired in sequence, and when two adjacent thyristors are conducting a commutating loop is formed. Into this loop must be injected a commutating voltage to turn off the outgoing thyristor. The simplest way of providing this voltage is to use conventional interpoles, but then the problem arises that at standstill or very low speeds there is inadequate induced voltage. Suggestions have been made[20] for providing an external commutating voltage, but this together with the large number of thyristors already required makes the whole system costly and complex and gives it very little merit over other means of obtaining a variable-speed drive such as an inverter-fed induction motor. This is particularly so when it is borne in mind that the commutatorless d.c. machine still has to be fed from a thyristor-controlled voltage source to be a variable-speed motor since the thyristors that do the commutation cannot at the same time be used for voltage control.

Some commutatorless d.c. machines have been produced for very high speeds of rotation and speeds of up to 100,000 rev/min have been claimed. These machines use a solid rotor in which the field coils are not rotating and a synchronous motor with this form of construction, or an induction motor are the only feasible ways of attaining such high speeds. Thyristors are not always used in these special applications; at lower current levels power transistors provide simpler switching circuitry. High speed machines of this kind may also require contactless switching of the thyristors or transistors from the shaft position and it is only the availability of optical systems using photo-transistors, or allied devices based on Hall probe detectors that have made the schemes possible. The semiconductor content of such machines may thus be said to be extremely high and varied.

8.3 *Thyristor-assisted Commutation*

As a half-way house to commutatorless d.c. machines and as a means of overcoming the commutation limitations of d.c. machines with difficult combinations of power and speed, a scheme has been proposed that uses thyristors to improve sliding-contact commutation rather than do away with it altogether.[21] The essential principle is to use the commutator only as an off-load switching device to distribute the armature tapping-points in sequence to an external pair of thyristors that continuously switch back and forth and so perform the switching process involved in commutation. In this way the number of thyristors involved is kept economically small and at low speeds commutation is still possible because the interruption of current will then be by these brushes even though the interpole induced voltage at these speeds is inadequate. Thus the major disadvantage of the commutatorless machine are overcome.

Such a scheme is shown on Fig. 9. The thyristors are fired alternately and while two thyristors are conducting, a closed commutating loop is formed into which a commutating voltage (normally by interpole induction) is injected to cause turn-off of the outgoing thyristor before a brush reaches the end of its segment.

The strength of this idea lies in the very minor way in which the machine system differs from a conventional d.c. machine: the field, armature and commutator constructions being essentially conventional and the additional thyristor equipment relatively uncomplicated and external.

Several variations of the principle are possible, sometimes using diodes instead of thyristors, or a combination of both. The essential requirement in all cases is to delay the turn-on of current until a brush is fully on a segment using a thyristor triggered at the appropriate moment, and relying on the turn-off and reverse-biasing of a thyristor or diode in series with the last brush to leave a segment.

Once again the impact on machine engineering has been wider than the original concept would suggest. Such machines require brushes that will accept rapid on/off pulses of current and extensive studies have been made in order to ensure satisfactory current collection under such unusual conditions. These studies have thrown new light on current collection problems in general, once more an illustration of how the introduction of a new device such as the thyristor has caused further thought on well established matters.

9 Other Applications

Space is not available to cover all the possible applications and apologies are made to innovators whose pet schemes do not seem to have been covered. A few of these may be mentioned very briefly. Instead of putting only rotating diodes on a brushless alternator thyristors are now used to give control of rotor-excitation without the time delay associated with field winding control.[22] Vibrating-contact voltage regulators and carbon-pile regulators for generator voltage control have been replaced by thyristors,[23] to gain more rapid response

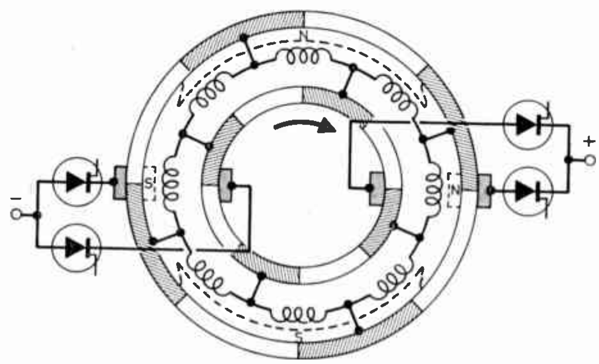


Fig. 9. Thyristor-assisted commutation.

and in most cases longer maintenance-free life. Perhaps some reference should be made to the mass-market devices such as power tools and domestic appliances. Although low-power in the eyes of power engineers, the simple application of naturally-commutated thyristors as speed controllers has been widespread and very effective.[24] Thyristors have in some cases been used as simple switches to replace contactors,[25] thus making such circuitry as is involved in, say, lift controls completely solid-state right through from the low-power logic circuits to the switching on and off of the motors themselves. Proposals have been made to use thyristors in on-loading tap-changing of transformers,[26] once more making full use of the simple natural-commutating ability of the devices in a closed loop in which an a.c. voltage exists. Theatre lighting has benefited greatly by replacing choke control by thyristors, reducing considerably the extent of the power cabling and the size of control consoles. The list is not endless, no list can be, but it is obvious that the number of possible applications is very great.

Emphasis has been placed on the thyristor as the controlled-semiconductor switching device, because its available rating far exceeds that of any other semiconductor device. Transistors can of course also be used as switches and, provided that they are used fully on or fully off, can perform at low power levels all the functions previously described. What cannot be utilized to any great extent is the progressive change in resistance possible with a transistor because, as has been emphasized several times, it would then become an excessively dissipating device for power use. At low power levels transistor switching can be used to advantage to eliminate commutating capacitors from the circuits involving forced-commutation and there are many applications in which transistors are used for armature current control of servo-motors.[27] However, most of the basic principles described in terms of thyristors are the same if transistors are used instead. Another power semiconductor device is the triac[28] which is equivalent to two back-to-back thyristors in one device but with only one trigger lead, and for some a.c. applications this device is to be preferred to the thyristor, the control of theatre lighting being a typical example.

10 Conclusions

Thyristors are currently available ex-stock in ratings up to 1200 V and 1000 A. This is indeed a power device and the impact of the thyristor on power engineering must continue to be very great. The stage has been reached where the better schemes have been economically assessed and some universally accepted priorities now exist for continued development. It is not necessarily the cleverest scheme that ultimately proves most useful. For a thyristor system to be really successful it must be commercially worthwhile and in this context it must not be forgotten that electrical machines are well developed and proven devices, incorporating in their development extensive know-how and operating experience. The extensive time-scale in developing a new machine system should also not be forgotten; electrical machines are not made and tested overnight and several years work may well be needed to bring to fruition any new method of machine construction and/or control.

A good example of the detailed care necessary in evaluating a successful thyristor scheme is in the use of thyristor convertors fed from polyphase a.c. supplies to control d.c. motors. A few years ago the cost of thyristors made it economic to use only half-wave convertor systems. This meant that the ripple content of the voltage applied to the motor was high and a limit occurred in the maximum motor size at which the system could be applied due to commutation problems, as the incorporation of fully-laminated field systems, added to the cost of the thyristors, did not make the overall arrangement economic above a certain size compared with a conventional Ward-Leonard system. With falling thyristor prices it became possible to use full-wave thyristor bridges and the reduced ripple content then eased the commutation problem and enabled the economic rating limit to rise to a sufficient extent to bring it into an area where much larger markets existed for such drives. This is also another good illustration of the close interaction between the thyristors and the machines to which they are applied.

The maintenance aspect must also be remembered. Electrical machines in industry have to be maintained and the men employed for such maintenance will not currently have many electronic skills. Psychology also has a bearing on the matter. There is the story of the maintenance engineer in a steelworks who could quite happily go home having had a last look at his Ward-Leonard generators noisily rotating, fully confident that they would still be turning when he returned the next morning. But he had no peace of mind when he had a look at his quiet static thyristor cabinet doing the same job and thought of hundreds of amps. charging back and forth through little thyristors, controlled by a complicated array of transistors and other electronic components. If the same evening his television set stopped working at home he would be unlikely to get any sleep at all! These men often have a say in what equipment should be installed in any future scheme and due understanding should be given to their reluctance sometimes to be persuaded from the conventional and well understood in favour of the (so they are told) 'modern' way.

Power engineers are learning to use the thyristor, but as has been said already, an integrated approach to the whole system is necessary for success. The power engineer must know sufficient electronics to appreciate the functioning of the whole system if the right systems are to be developed. As the attitudes of mind in power engineering are often so different from those in electronic engineering it will be no mean task to marry the two and produce engineers who have the necessary broad versatility and confidence. It is a challenge to engineering education if nothing else.

11 References

1. Proceedings of the British Railways Electrification Conference (London, 1960).
2. Ford, A. W., 'Brushless generators for aircraft—a review of current developments', *Proc. Instn Elect. Engrs*, **109**, No. 1, pp. 437–55, January 1962.
3. Stephen, D. D., 'Industrial applications for brushless synchronous motors', I.E.E. Conference Publication No. 10, pp. 167–76, 1965.
4. Pelly, B. R., 'Thyristor Phase-Controlled Convertors and Cycloconverters', p. 10 (Wiley-Interscience, New York, 1971).
5. Schaefer, J., 'Rectifier Circuits', pp. 183–91 (Wiley, New York, 1965).
6. Schofield, J. R. G., Smith, G. A. and Whitmore, M. G., 'The application of thyristors to the control of d.c. machines', I.E.E. Conference Publication No. 17, pp. 219–29, 1965.
7. Bedford, B. D. and Hoft, R. G., 'Principles of Inverter Circuits', pp. 316–57 (Wiley, New York, 1964).
8. Ibid. pp. 336–7.
9. Adkins, B. and Gibbs, W. J., 'Polyphase Commutator Motors' (Cambridge University Press, 1951).
10. Bedford, B. D. and Hoft, R. G., loc. cit., pp. 89–315.
11. Bradley, D. A., *et al.*, 'Adjustable-frequency invertors and their application to variable speed drives', *Proc. Instn Elect. Engrs*, **111**, No. 11, pp. 1833–46, November 1964.
12. Davis, E. W. and Meier, R. W., 'Application of inverter drives', *IEEE Trans. on Industry and General Applications*, **IGA-5**, No. 1, pp. 45–52, January/February 1969.

13. Mokrytzki, B., 'Pulse width modulated inverters for a.c. motor drives', *IEEE Trans. on Industry and General Applications*, **IGA-3**, No. 6, pp. 493–503, November/December 1967.
14. Bowler, P. and Shelley, R. B., 'Accurate control of a high speed drive', I.E.E. Conference Publication No. 93, pp. 229–231, October 1972.
15. Rissik, H., 'Mercury-Arc Current Convertors', pp. 237–56, 2nd Edition (Pitman, London, 1941).
16. Slabiak, W. and Lawson, L. J., 'Precise control of a three-phase squirrel-cage induction motor using a practical cycloconvertor', *IEEE Trans. on Industry and General Applications*, IGA-2, No. 4, pp. 274–280, July/August 1966.
17. Piper, R. L. and Wilcock, G. W., 'Control Aspects of Naturally-Commutated Cycloconvertors for Aircraft Use', R.A.E. Technical Report, November 1972.
18. Tam, C., 'General survey of hvdc projects', *Direct Current* (New Series), **1**, No. 4, pp. 128–36, 1970.
19. Dewey, C. G., Dranchak, M. A. and Leete, B. D., 'A 200 kV solid-state valve for hvdc systems', I.E.E. Conference Publication No. 53, pp. 372–80, May 1969.
20. Tustin, A., 'Electric motors in which commutation is by switching devices such as controlled silicon rectifiers', *Proc. Instn Elect. Engrs*, **111**, No. 1, pp. 151–2, January 1964.
21. Bates, J. J. and Sridhar, T. V., 'Thyristor-assisted sliding contact commutation', *Proc. Instn Elect. Engrs*, **113**, No. 2, pp. 339–41, February 1966.
22. Aked, C. K. and Smith, A. W., 'QUAVER—a new concept in automatic voltage control', *J. R. Naval Sci. Serv.*, **23**, No. 4, pp. 219–31, July 1969.
23. Gilchrist, D. C., 'Thyristor automatic voltage regulators and static exciters', *G.E.C. J.*, **35**, No. 1, pp. 3–13, 1968.
24. Collie, A. A., 'Solid-state control of domestic appliances', *Electronics & Pwr*, **18**, pp. 19–22, January 1972.
25. Feltbower, B., 'Thyristor contactors', I.E.E. Conference Publication No. 53, pp. 448–64, May 1969.
26. Roberts, M. E. and Ashman, W. G., 'A thyristor-assisted mechanical on-load tap changer', ibid., pp. 185–92.
27. Watston, J. A. and Miller, J. R., (Ed.), 'Transistor Circuit Design', McGraw-Hill (1963), pp. 238–50 and 481–83.
28. Cooper, D., 'Bi-directional thyristors for a.c. power control', *Elect. Rev.*, pp. 718–20. November 1968.

Manuscript received by the Institution on 8th November 1972. (Paper No. 1503/IC70.)

UDC 616-089.2:621.3.049.77

Microelectronic devices for surgical implantation

P. E. K. DONALDSON,
M.A., C.Eng., M.I.E.E.*

and

J. G. DAVIES,
M.B.E.†

SUMMARY

The surgical implantation of an electronic device is shown to have been made possible by advances (1) in the understanding of the stimulation processes of nerves, (2) in the development of electrical recording of nerve and muscle activity, (3) in transistor technology, (4) in implant technology. Significant developments in these areas are identified, leading up to the history of their bringing together, during the past decade, to enable successful cardiac pacemakers (neuro-electroneural prostheses) to be produced. Recent developments in electroneural systems for visual prosthesis, a pain suppressor and an incontinence aid, and in neuroelectric devices for muscle control are described briefly.

* *MRC Neurological Prostheses Unit, Institute of Psychiatry, De Crespigny Park, London S.E.5*

† *Cardiac Department, St. George's Hospital Medical School, Hyde Park Corner, London S.W.1*

1 Introduction

Of the ways in which electronic engineering is used in support of the practice of Medicine, it is probably true to say that most are in the field of clinical measurement. Such measurements may be termed non-invasive, i.e. no surgery is required, as when, for example, ultrasonic echo methods are used to locate structures deep within a mass of tissue; or when, by scanning the body with a probe which responds to gamma radiation, one builds up a map of the disposition of some radioactive tracer; or when one takes an electrocardiogram. Alternatively, some surgical procedure may be necessary, as when a catheter is passed into and along a blood vessel and so into the interior of the heart, enabling pressure, flow, oxygen tension, etc. determinations to be made at the exact spot required. In designing equipment to perform such measurements it may be convenient to implement the electronic part with transistors, but it is seldom essential to do so. Such equipment could be implemented using thermionic valves, and in the past, of course, necessarily was.

There is, however, another class of medically-used electronic device whose existence is totally dependent on the transistor. This is the microelectronic implant, placed inside the body by surgical operation, and left there for an indefinite period. The best known of these is the cardiac pacemaker, and it is the purpose of this paper to consider the pacemaker and some allied devices which, if less well known, nevertheless are likely to grow in importance in the near future.

When Chardack and his colleagues announced the implantable fixed-rate pacemaker in 1960,[1] it was clear that here was an important development, but it was perhaps not clear at the time just how important, how seminal a development it was. Here is a commercially manufacturable device which enables many people, who would not otherwise be able to do so, to live substantially normal lives. But it occupies also an important position in the historical evolution of its particular branch of electromedical engineering. For, with its more elaborate derivatives, it represents a confluence of four streams of development: the two electrophysiological techniques of recording from, and stimulating, nerve or muscle; the technology of the transistor, and of the small passive components to go with it; and what may be called implant technology, the art of making things which may be put into the body, so that neither is damaged by the other. Furthermore, the pacemaker stands at the point of departure for a whole class of new devices which go under the general heading of *neurological prostheses*. A prosthesis is an artificial device which replaces some part of the body which is absent or inoperative, e.g. false teeth or an artificial hip-joint. A neurological prosthesis is thus an artifact which aims to replace missing or defective nervous tissue, or structures physiologically allied to nervous tissue (Fig. 1).

2 Historical Development

2.1 *Stimulation*

That electricity can cause stimulation of nerve and contraction of muscle has been known for nearly 200

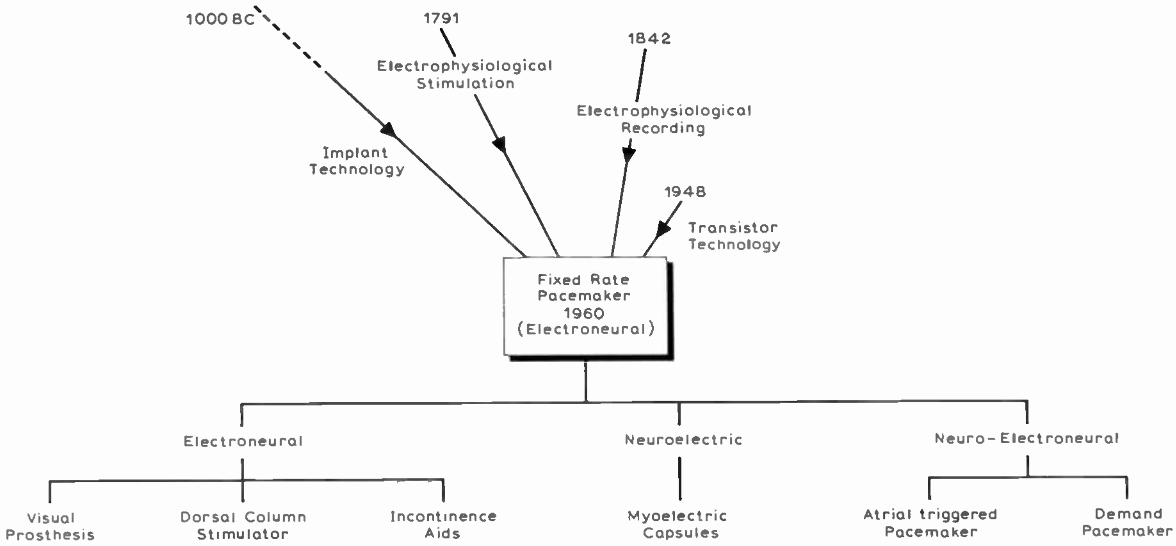


Fig. 1. The evolution of microelectronic implants

years. Galvani's famous observations with the legs of frogs were published in 1791; he claimed that stimulation could be produced by an electric discharge, either atmospheric or from an electric machine, or by touching the nerve with a couple composed of dissimilar metals. Although, in the latter case, there was subsequently a good deal of acrimony between the followers of Galvani and those of Volta as to precisely why stimulation occurred, there was no doubt that, here also, the effect was electrical.[2]

We now know a great deal about the biophysics of stimulation; but it has been known since the end of the last century that stimulation will occur at a particular point if the current there is in the correct direction, exceeds a certain minimum strength, does not grow too slowly, and exceeds a certain duration. Typical stimulus durations in electrophysiological work lie in the range 0·1–1·0 ms.

2.2 *Electrical Recording of Nerve and Muscle Activity*

The story here may be traced back to 1746, and the invention by van Musschenbroeck of the Leyden jar.[2] Five years later, a traveller in West Africa, Michel Adanson, received a shock from an electric catfish. He said 'Son effet ne m'a pas paru différer sensiblement de la commotion électrique de l'expérience de Leyde'. This seems to have been the first occasion in which it was recognized that at least some biological tissues generate electricity.[3] Demonstration of the much feebler electric effects set up by contracting muscle did not come till 1842, when Matteucci devised the 'rheoscopic frog'; if the nerve of frog nerve-muscle preparation A is laid across the muscle of a similar preparation B, and the nerve of preparation B is then stimulated by pinching it, both muscles contract.[2] Evidently nerve A is stimulated by something coming from muscle B; since the two muscles contract practically simultaneously, the effect can scarcely be chemical and must therefore be electrical. Matteucci was also able to show an electrical change when a nerve was caused to carry whatever it was that nerves do carry. He managed this by demonstrating—using the recently invented galvonometer—the diminution in the so-called 'injury current' of a large nerve, comprising many nerve fibres, when the nerve was repeatedly stimulated.[2] In 1850, von Helmholtz was able to measure the speed at which the nervous message is carried (30 m/s for the nerve he investigated) and in the 1860s Bernstein, in a beautiful example of the early use of sampling technique, was able to show, using the relatively sluggish galvanometers of the time, that the nervous message consisted of a volley of electrical disturbances, which travelled along without attenuation, each setting up as it passed the recording electrodes an 'action potential' of duration of the order of a millisecond.[2] Each such disturbance in each nerve fibre represents a quantum of nervous information. Demonstration of action potentials in real time, as a wave on the screen of a cathode-ray tube, was achieved by Gasser and Erlanger, using a three-stage valve amplifier in 1922.[4]

More recently there has been a great burgeoning of electrophysiological recording using electronic techniques, both in the service of scientific knowledge and as an aid to clinical practice. We may note two out of the many ways in which such recording has developed; they are interesting because they are so different. One is the demonstration by Berger, in 1929, of the electroencephalogram or 'brainwave'.[5] In this technique, rather large electrodes are attached to the subject's scalp and connected to very high gain amplifiers and recorders. The records so obtained reflect the activity of huge numbers of cells in the part of the brain beneath the electrodes. The method is now a standard clinical tool.

The other arises from the development by Ling and Gerard, described in 1949,[6] of a method of making a saline-filled, glass micro-pipette having a tip diameter of less than 1 μm. This provides a very fine electrode which may be pushed inside a muscle fibre, nerve fibre or nerve cell, enabling the biophysics of just that one fibre or cell to be investigated. These studies have largely confirmed the predictions made by Bernstein 50 years before, that the inside of a resting fibre or cell is maintained electrically negative with respect to the outside, and that the passage of the disturbance which gives rise to the action potential is marked by a transient disappearance of this internal negativity; in point of fact, the interior is found to become, momentarily, some 30 mV *positive* before reverting to the resting level of about 80 mV negative.

2.3 *Transistor Technology*

Biologists were not slow to appreciate the qualities of ruggedness, small size and economy of power consumption offered by the transistor. An early application of transistors to physiology was that of French,[7] who in 1953 described a small cat-borne radio receiver for remotely-controlled stimulation of part of the brain of an unanaesthetized, unrestrained animal. Use of transistors for electrophysiological recording came a year or two later.[8–10] By 1958, Beenken and Dunn[11] were able to describe a miniature radiotelemetry system for up to 10 physiological variables, to be worn by a human subject, having a range of 30 metres and weighing about 1 kg.

2.4 *Implant Technology*

The use of foreign materials for prosthetic purposes in or about the body is of great antiquity. The Etruscans fashioned false teeth from the teeth of oxen, attaching them in the mouths of patients by gold bands which hooked over neighbouring natural teeth. Dental fillings of a mixture of mastic and alum were described by the Persian physician Rhazes (865–925) and gold fillings by Giovanni of Arcola toward the end of the fifteenth century.[12] The modern amalgam filling is a nineteenth century development. Even older is the choice of materials for sewing and tying in surgery. The textbook ascribed to Susruta (about 1000 BC) recommends cotton, horsehair and animal tendons for sutures,[13] materials in use today. Antyllus, in the 3rd century AD, mentions the use of linen thread and catgut,[14] again materials in current use. Robert Liston's 'Practical Surgery', published in London in 1846, talks of 'ligatures of wire (a composition of platina etc.) possessing great ductility and strength' but warns that such ligatures have no advantage over silk or thread. There is in the library of King's College Hospital Medical School a sample of silver wire as used by Lord Lister (1827–1912) for surgical work. Magnesium wire was tried by Huse in 1878,[14] in search of material that was strong but would eventually be absorbed by the body.

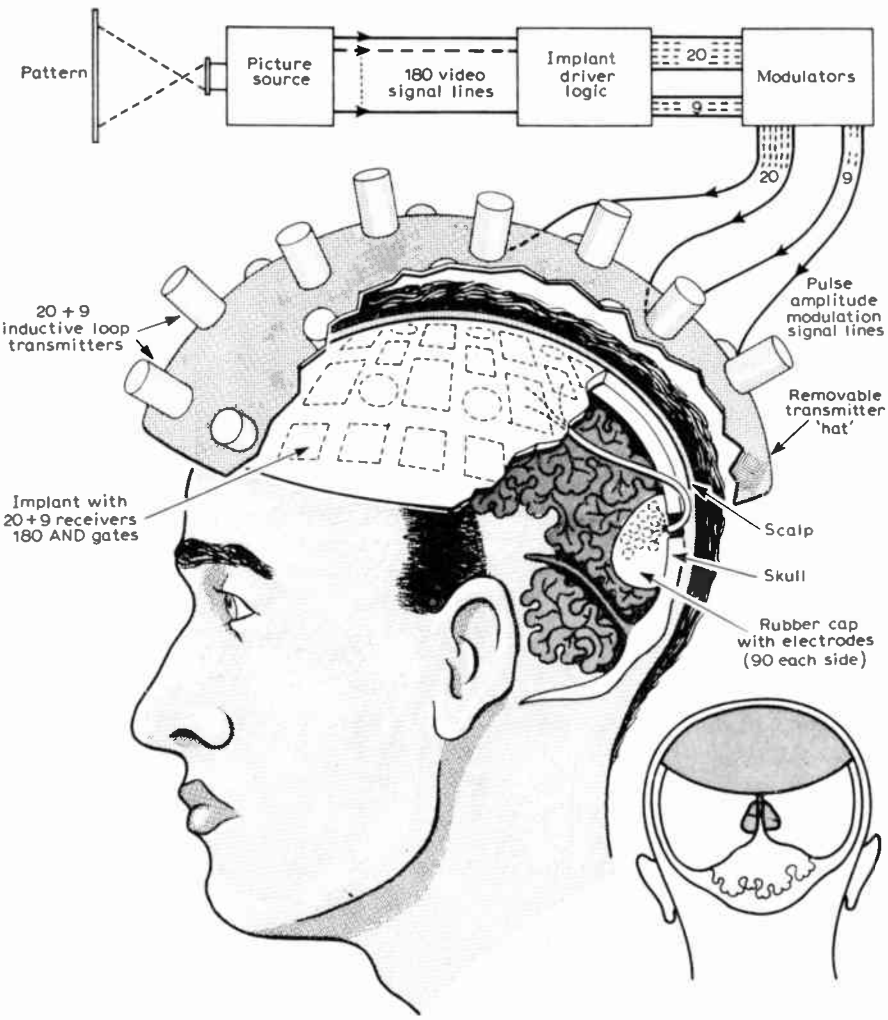
In dealing with fractures of the limb, the traditional treatment has been to immobilize the limb using either splints or gypsum plaster, so as to constrain the broken ends of bone to remain in suitable apposition to one another while healing takes place. Clearly a more rigid constraint is applied if the constraining apparatus connects the bones themselves, rather than the surrounding soft tissues. In 1870, L. J. B. Berenger-Ferand described, but did not claim as original, the fixation of fractures by driving screws through a screwplate on the *outside* of the limb, down through the soft tissues of the limb and into the underlying bone.[14] It is difficult to believe that patients took kindly to this procedure. The modern arrangement, in which the bones are held by a screwplate placed directly against them, was pioneered by Sir William Arbuthnot Lane and by the Belgian, A. Lambotte. Lane's book, published in 1914, shows many x-rays of his screwplates and screws in position;[15] these are genuine implants, in a sense in which dental fillings and sutures perhaps are not. The materials used by Lane would not now be considered ideal; he described his screws as 'the ordinary wood screws of the trade' and it is not to be wondered that his implants encountered tissue reactions and corrosion. More suitable materials, e.g. stainless steels, were investigated during the 1920s, and in 1938 Venable and Stuck published a famous paper[16] in which the contribution of electrolysis to corrosion was recognized. The authors drove pairs of screws, of different metals, into the bones of experimental animals, subsequently connecting them with a microammeter. Current always flowed between dissimilar screws unless one of the pair was of the Co-Cr-Mo alloy, Vitallium. They concluded that the passivity of the Vitallium surface must be unaffected by tissue fluid and that therefore, taking also into account some excellent clinical results they were getting using Vitallium screws and plates, this material must be considered outstanding for implant construction.

A parallel history is that of the fixation of the broken heads of femurs. Langenbeck sought to repair such a fracture by nailing in 1850, and various similar attempts were made up to the time of the first world war.[14] Among the difficulties which beset these early efforts were poor mechanical design of the nail, unsuitable materials and, no doubt, sepsis. In the post-war period, improvements were gradually made, and the Smith-Petersen nail, described in 1931, marks an important advance.[17] The design of this is such that the volume is small (so that not too much bone has to be removed) but the surface area is large (so that the nail is tightly gripped). The material used is described by the author as 'rustless steel'.

Neurosurgeons have been responsible for a good deal of innovation in implantation. In the repair of damage to the skull, for example, artificial materials may be used to bridge over or fill in gaps left after the removal of broken bone fragments. Fraenkel described the use of celluloid for this purpose in dogs, in 1890,[18] but refers in his paper to even earlier workers using cork and rubber. According to Gurdjian and Thomas,[19] other materials which have been used, at one time or another, for skull repair are aluminium, gold, silver, platinum, steel, stainless steel, Vitallium and lead. Some of these would have been more successful than others. Favoured modern materials are tantalum, titanium and methyl methacrylate. Tantalum is popular because, like Vitallium, it is always coated with a tough passive layer. Polyethylene has also been used.[20]

2.5 *Implanting Electrical Devices*

In designing electrical equipment for implantation into human or animal tissue, the motive is usually to avoid having wires passing through the skin. If one wishes to stimulate some deep structure, let us say, it is possible to take wires from an external stimulating pulse generator,



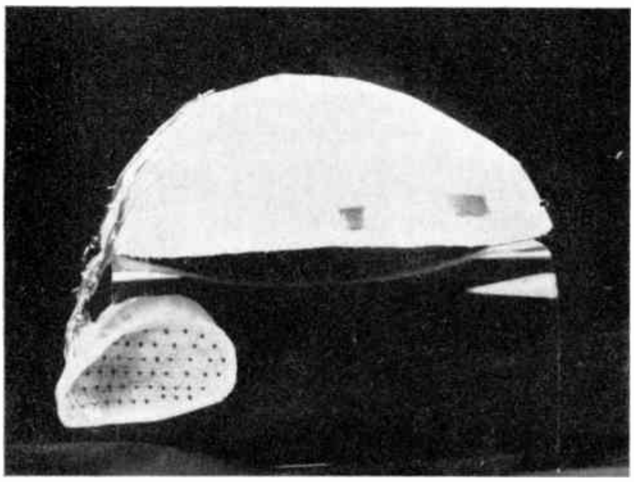


Fig. 6. Principle of proposed 180-point visual prosthesis system. The visual cortex of the brain is stimulated by electrodes fed from microelectronic receivers in an implant between the skull and the scalp. In practice the transmitter 'hat' is in contact with the head. The implant could not at that time be made small enough and a 75-point scheme was implemented instead. This successful 75-point implant is shown on the left.

(Diagram reproduced from the article 'Artificial vision' in *Wireless World*, May 1971, by courtesy of the Editor).

4 Progeny of the Pacemaker: the Neurological Prostheses

Neurological prostheses can be categorized by the nature of the interaction or interactions they make with biological material. Thus a fixed rate pacemaker is *electroneural*, because the direction of information flow at the interface is from electronic device to nervous (actually muscle) tissue. A *neuroelectric* prosthesis is one in which information flows the other way (see below). A demand pacemaker is neuro–electroneural, because information flows from tissue to device and back from device to tissue again. Electro–neuroelectric and more elaborate prostheses do not yet exist, except as conceptual possibilities.

We look now at some neurological prostheses, other than pacemakers, which have appeared since 1960.

4.1 *Electroneural Prostheses*

A visual prosthesis is one intended to provide a blind person with some capacity for seeing. The quality of vision obtained may not be very high, but quite poor vision is better than none at all. We are concerned here with patients blinded by damage to the eye or optic nerve. The idea is that, since the retina of the healthy eye is mapped by nervous connexions one-to-one on to the surface of the visual cortex of the brain, it might be possible to evoke the perception of visual images by spatially patterned stimulation of the cortical surface from an array of chronically-indwelling electrodes. Since the number of electrodes, in early models at least, will be of the order of hundreds, whereas the number of optic nerve fibres is of the order of millions, the crudity of artificial vision of this kind can be judged. Stimulating current would be supplied from an implanted microelectronic device, under the scalp, that would receive radio signals from a number of oscillators mounted inside a special hat, put on when the prosthesis is required. The oscillators would, in turn, receive signals from a miniature photo-electric camera by way of suitable logic circuitry. The first visual prosthesis was described by Brindley and Lewin in 1968: it showed great promise.[34] Stimulation of a particular place on the cortical surface produced the perception of a point of light, fixed in the visual field. Stimulation of two or more places at the same time produced two or more white points, in the expected relative position. Further engineering development was put in hand, and a new prosthesis was implanted into a second patient in February 1972; this is also giving encouraging results.[35, 36] (Fig. 6). The prospects for future, improved visual prostheses seem excellent.

In 1965, Melzack and Wall[37] described a novel theory of the physiological mechanism of pain. They suggested that if, at the point on the dorsal side of the spinal cord where nerve fibres mediating a particular pain enter it, the large diameter fibres could be caused to 'fire' at a higher rate, or the small fibres at a lower, then the intensity of pain experienced should be diminished. Since the current required to stimulate a large fibre is lower than that required for a small one, it should be possible to increase the activity of the large fibres artificially by the use of an implanted stimulator whose output is adjusted to just the right strength. In patients suffering from chronic intractable pain, use of such a stimulator should, if Melzack and Wall are right and the stimulator is correctly fitted, cause the pain to be reduced or abolished. In the United States a number of attempts have been made to carry out this procedure,[38, 39] and dorsal column stimulators are commercially available. The technique has been tried in this country, but without much success.[40]

In the treatment of urinary incontinence, it is often found helpful to improve the tone of the urethral sphincter and the pelvic floor by the use of electrical stimulation. To do this, a power of about 100 mW is continuously required. Therefore, where an implanted stimulator is used, this has customarily taken the form of a simple secondary coil which is excited from an external primary winding. Commercially manufactured coils are available for this purpose. Recently, Brindley has suggested[41] that it might be possible to achieve control of both bladder emptying and sphincter closure by suitably patterned stimulation of the relevant spinal roots, where motor nerves leave the spinal cord *en route* for their respective muscles. The source of stimulation would be an active microelectronic implant complete with batteries. An advantage of stimulating nerve, rather than muscle, is that it requires less power to do it; hence a self-powered nerve stimulator can be either smaller, or will run longer before the batteries have to be changed or recharged. The utility of motor root stimulation for bladder and sphincter control is at present under evaluation in baboons.

4.2 *Neuroelectric Prostheses*

Persons who have had an arm amputated at the elbow will in general retain the muscles in the upper arm which originally flexed and extended the elbow joint; they can still contract these muscles at will, although of course no useful purpose is normally served thereby. In 1955, A. Nightingale and his colleagues at Guy's Hospital showed[42, 43] that it was feasible to lead action potentials from these muscles, recorded via electrodes on the surface of the arm, to amplifier-rectifier and filter sets, thereby producing electric signals which could be made to fulfil some useful purpose, such as opening and closing the hook on the end of an arm prosthesis. One difficulty with this idea is that it is rather inconvenient to have to wear electrodes stuck to the skin, and another is that action potentials recorded on the surface are not very large, and are in consequence liable to be interfered with by extraneous signals. Much bigger action currents are present within the actual muscle itself, and a way to solve both difficulties is to implant into each muscle a radio-telemetering capsule (myo-electric capsule) which makes available to equipment outside the body a signal corresponding to the action currents flowing in the tissue immediately surrounding the capsule. Such radio-telemeters can be made extremely small, about the size of half a matchstick; they may work in various ways. In the capsule described by the Swedish workers Hirsch, Kadefors, Kaiser and Petersen,[44] the capsule collects and rectifies r.f. energy transmitted to it, from outside the body, at 3 MHz, and uses this to power a 450 kHz Hartley oscillator which is frequency modulated by the muscle action current. The output of the oscillator is picked up in the normal way by an f.m. receiver outside the body. The capsule contains two coils, two varactors and two transistors and it occupies less than $0{\cdot}3\ \mathrm{cm}^3$. The EMGOR device of R. E. Reilly is even smaller and simpler.[45] It contains only a small ferrite-cored inductor tuned by a pair of varactors whose bias is the muscle action current. Fluctuations in the action current are detected as changes in the impedance reflected into a coil, outside the body, which is magnetically coupled to the EMGOR coil.

Although patients can make such myoelectric capsules work, it has to be admitted that, at present, most powered artificial limbs remain mechanically controlled.

5 Conclusions

The foregoing review has been necessarily brief and is certainly incomplete. We hope, however, that we have shown that the placing of electronic devices semipermanently within the body is now a considerable activity with roots buried deep in the past and with an exciting future; if the transistor is the junior partner in the four technologies from which the neurological prostheses derive, its impact is none the less for that.

6 References

1. Chardack, W., Gage, A. and Greatbatch, W. A., 'A transistorized self-contained implantable pacemaker for the long-term correction of complete heart block', *Surgery*, **48**, pp. 643–54, 1960.
2. Brazier, M. A. B., 'Handbook of Physiology', Section 1, 'Neurophysiology', pp. 17–23. (American Physiological Society, Washington D.C., 1959.)
3. Keynes, R. D., 'The generation of electricity by fishes', *Endeavour*, **15**, pp. 215–22, 1956.
4. Gasser, H. S. and Erlanger, J., 'A study of the action currents of nerve with the cathode ray oscillograph', *Amer. J. Physiol.*, **62**, pp. 496–524, 1922.
5. Berger, H., 'Uber das Elektrenkephalogramm des Menschen', *Arch. Psychiat.*, **87**, p. 527, 1929.
6. Ling, G. and Gerard, R. W., 'The normal membrane potential of frog satorius fibres', *J. Cell Comp. Physiol.*, **34**, pp. 383–96, 1949.
7. French, J. D., 'Transistor circuits in remote stimulation', *EEG Clin. Neurophysiol.*, **5**, pp. 613–6, 1953.
8. Bates, J. A. V. and Cooper, J. D., 'Problems in the design of a 21-channel eeg machine', *EEG Clin. Neurophysiol.*, **8**, pp. 703–4, 1956.
9. George, R. E., 'Transistor preamplifiers for electromyography', *Proc. Electrophysiol. Tech. Assoc.*, **7**, pp. 6–7, 1956.
10. MacNichol, E., 'Transistors in physiological amplifiers', *Trans. IRE Medical Electronics*, No. PGME12, pp. 15–24, 1958.
11. Beenken, H. G. and Dunn, F. L., 'Short distance telemetering of physiological information', *Trans. IRE on Medical Electronics*, No. PGME12, pp. 53–7, 1958.
12. Lufkin, A. W., 'History of Dentistry', 2nd ed. (Kimpton, London, 1948.)
13. Melle, G. J., 'Early history of the ligature', *S. Afr. Med. J.*, **8**, pp. 290–2, 1934.
14. Bick, G. M., 'Source Book of Orthopedics', 2nd ed. (Williams and Wilkins, Baltimore, 1948.)
15. Lane, W. A., 'Operative Treatment of Fractures', 2nd ed. (Medical Publishing Company, London, 1914.)
16. Venable, C. S. and Stuck, W. G., 'Electrolysis controlling factor in the use of metals in treating fractures', *J. Amer. Med. Ass.*, **111**, pp. 1349–52, 1938.
17. Smith-Petersen, M. M., Cave, E. F. and Vangorder, G. W., 'Intracapsular fractures of the head of the femur', *Arch. Surg.*, **23**, pp. 715–59, 1931.
18. Fränkel, A., 'Uber deckung von Trapanationsdefekten am Schadel durch Heteroplastik', *Wien Klin. Wiss.*, **3**, pp. 475–6, 1890.
19. Gurdjian, E. S. and Thomas, L. M., 'Operative Neurosurgery', 3rd ed. (Williams and Wilkins, Baltimore, 1970.)
20. Alexander, E. and Dillard, P. H., 'Use of pure polythene plate for cranioplasty', *J. Neurosurg.*, **7**, pp. 492–8, 1950.
21. Loucks, R. B., 'Preliminary report of a technique for stimulation of tissues beneath the integument', *J. Comp. Psychol.*, **16**, pp. 439–44, 1933.
22. Fender, F. A., 'Convulsive phenomena produced by a new method of remote excitation', *Arch. Neurol. Psychiat. (Chicago)*, **45**, pp. 617–32, 1941.
23. Harris, G. W., 'The innervation and actions of the hypophysis, an investigation into the method of remote-controlled stimulation', *Phil. Trans. R. Soc.*, **232**, pp. 385–441, 1947.
24. Mackay, R. S. and Jacobson, B., 'Endoradiosonde', *Nature*, **179**, pp. 1230–1240, 1957.
25. Farrar, J. T., Zworykin, V. K. and Baum, J., 'Pressure-sensitive telemetry capsule for study of gastrointestinal motility', *Science*, **126**, pp. 975–6, 1957.
26. Jacobson, B., 'Endoradiosonde technique—a survey', *Med. Electronic Biol. Engng*, **1**, pp. 165–78, 1963.
27. Hyman, A. S., 'Resuscitation of the stopped heart by intracardiac therapy. Experimental use of an artificial pacemaker' *Arch. Int. Med.*, **50**, p. 283, 1932.
28. Zoll, P. M., 'Resuscitation of the heart in ventricular standstill by external electrical stimulation', *New Engl. J. Med.*, **247**, p. 768, 1952.
29. Davies, J. G., Leatham, A. and Robinson, B. F., 'Ventricular stimulation by catheter electrode', *Lancet*, **1**, pp. 583–4, 1959.
30. Davies, J. G., 'Artificial cardiac pacemakers for the long-term treatment of heart block', *J. Brit. Instn Radio Engrs*, **24**, pp. 453–6, December 1962.
31. Furman, S. and Robinson, G., 'The use of intracardial pacemakers in the correction of total heart block', *Surg. Forum*, **9**, pp. 245–8, 1958.
32. Reynolds, G., Personal communication.
33. Trotman, R. E., 'Implantable cardiac pacemakers' (Report of a colloquium), *The Radio and Electronic Engineer*, **42**, pp. 464–6, October 1972.
34. Brindley, G. S. and Lewin, W. S., 'The sensations produced by electrical stimulation of the visual cortex', *J. Physiol.*, **136**, pp. 479–93, 1968.
35. Brindley, G. S., Donaldson, P. E. K., Falconer, M. A. and Rushton, D. N., 'The extent of the region of occipital cortex that, when stimulated, gives phosphenes fixed in the visual field', *J. Physiol.*, **225**, pp. 57–58, 1972.
36. Donaldson, P. E. K., 'An experimental visual prosthesis', *Proc. Instn Elect. Engrs*, (in press).
37. Melzack, R. and Wall, P. D., 'Pain mechanisms: a new theory', *Science*, **150**, pp. 971–9, 1965.
38. Shealy, C. N., Mortimer, J. T. and Reswick, J. B., 'Electrical inhibition of pain by stimulation of the dorsal columns: preliminary clinical report', *Anaesthesia & Analgesia*, **46**, pp. 489–91, 1967.
39. Mortimer, J. T., 'Pain Suppression in Man by Dorsal Column Electro-analgesia', Engineering Design Centre, Report No. 4-68-21. Case Western Reserve University, Cleveland, Ohio.
40. Bunyard, P., 'Intractable pain—how treatable is it?', *World Medicine*, **7**, No. 6, pp. 17–25, 1st December 1971.
41. Brindley, G. S., 'Experiments directed towards a prosthesis which controls the bladder and the external sphincter from a single site of stimulation', Biological Engineering Society Conference on Materials, Prosthetics and Aids for the Handicapped, Liverpool, 1972.
42. Battye, C. K., Nightingale, A. and Whillis, J., 'The use of myo-electric currents in the operation of prostheses', *J. Bone Joint Surg.*, **37B**, pp. 506–10, 1955.
43. Bottomley, A. H., Kinnier-Wilson, A. B. and Nightingale, A., 'Muscle substitutes and myoelectric control', *The Radio and Electronic Engineer*, **26**, pp. 439–48, December 1963.
44. Hirsch, C., Kadefors, R., Kaiser, E. and Petersen, I.,' An implantable microcircuit for long term telemetry of myopotentials', 7th International Conference on Medical and Biological Engineering Stockholm, 1967.
45. Reilly, R. E., 'EMGOR; an implantable sensor for myoelectric signals', *Proc. Instn Mech. Engrs*, **183**, Pt. 3J, pp. 109–16, 1968–1969.

Manuscript first received by the Institution on 1st December 1972 and in revised form on 1st January 1973. (Paper No. 1504/MBE13.)

UDC 621.382: 621.317.7

The influence of semiconductors on the evolution of typical laboratory instruments

J. B. IZATT,
B.Sc., Ph.D.*

SUMMARY

The influence of the transistor on digital instruments is considered first and the example of its application to digital voltmeters shows how the product became practical in the late 1950s and then went through successive improvements in performance as the devices available improved. Later developments made the instrument more automatic and improved the cost and reliability. The influence of the transistor on analogue instruments did not start till much later but eventually the transistor displaced the valve and provided a much improved performance. Illustrations from oscilloscope design show the influence of tunnel diodes on trigger circuits and how the alternate series/shunt feedback configuration is used in amplifiers to give very wide bandwidths. It is generally true that the introduction of semiconductors has resulted in cheaper instruments with better performance doing more complex tasks.

* *The Solartron Electronic Group Ltd., Victoria Road, Farnborough, Hampshire.*

1 Introduction

Instrumentation is essential in all branches of engineering and therefore the influence of semiconductors is felt over a very wide field. However, in any one branch of engineering, instrumentation tends to be a relatively small part of the whole and consequently the instruments which can be developed at any one time are limited by the devices available for other purposes. This is in marked contrast to the situation in digital computers, for example, where the application is sufficiently large to merit special attention by the component manufacturers.

It is for this reason that the influence of the semiconductors can be divided into two broad categories:

(i) digital instruments

(ii) analogue instruments

The transistor has had its earliest and greatest effect on the first of these and therefore the first part of the paper deals with this aspect while the following part is devoted to analogue instrumentation.

It is thought that the various developments will be clearer if attention is restricted to only one or two types of instrument. Accordingly, digital voltmeters have been chosen to represent digital instrumentation and oscilloscopes to represent the analogue field. Neither of these instruments is strictly 'all-digital' or 'all-analogue' but the division is a convenient one for the present discussion. Other instruments will be mentioned where appropriate.

2 Digital Instrumentation

2.1 *Early Uses of the Transistor (The late 'fifties)*

As noted in the introduction, specific transistor types were not developed to meet instrumentation needs and therefore the instrument designs had to be adapted to the devices available. The first widely available transistors were the germanium alloy type, the OC71 and OC72 being typical examples. These were suitable for low-speed logic and could be used as relay drivers because of the relatively good bottoming performance.

In the late 1950s, digital voltmeters were just being developed and as most of the designs involved relays, logic or switching of some sort, transistors were among the possible devices which could be used. In essence a digital voltmeter may be considered to consist of an input amplifier, a reference source, means for counting, and a detector for comparing the input voltage with the voltage generated internally. Amongst designers there was little unanimity on the methods of counting and logic control and early designs used uniselectors, dekatrons and cold cathode tubes as well as relays.

Figure 1 shows one of the first digital voltmeters using transistors. In this design OC77s were used to drive relays, the input voltage being measured by successive approximation. A valve was used for the input stage to maintain a high impedance but otherwise transistors were used and considerable effort was devoted to making a small instrument. The overall dimensions are 16·5 cm high × 19·6 cm wide × 30·5 cm deep ($6\frac{1}{2} \times 7\frac{3}{4} \times 12$ in).

Fig. 1. Digital voltmeter from the late 'fifties.

The next step in development was to replace the relays by transistors. Figure 2 shows the type of circuit used and illustrates some of the difficulties confronting the instrument designer during the late 'fifties and 'sixties. TR3 and TR4 act as a switch which determines whether or not current flows in the reference resistor R (the value of which determines the size of digit generated). The reference voltage was switched from +6 V to −6 V for different polarity voltmeter inputs and therefore large drive voltages were necessary from bistable TR1 and TR2. The transistor used for all four positions was the TK40C which had the remarkable base-emitter breakdown rating of 40 V. Two transistors were necessary for the switch because the leakage in the off position was too great. R1 and MR1 were intended to equalize the current drawn from the reference rail because TR3 base current flows from the reference source when it is on but not when it is off. A further complication arose from the high current that was necessary to achieve a reasonable switching speed in TR1 and TR2. The dissipation in the collector load is 900 mW when the transistor is on and therefore it was necessary in those days to use wire-wound resistors. Some of these resistors proved to be very unreliable because there was insufficient dissipation in the resistor to drive off any moisture and the wire corroded. Components have now been developed to match the performance of the transistor.

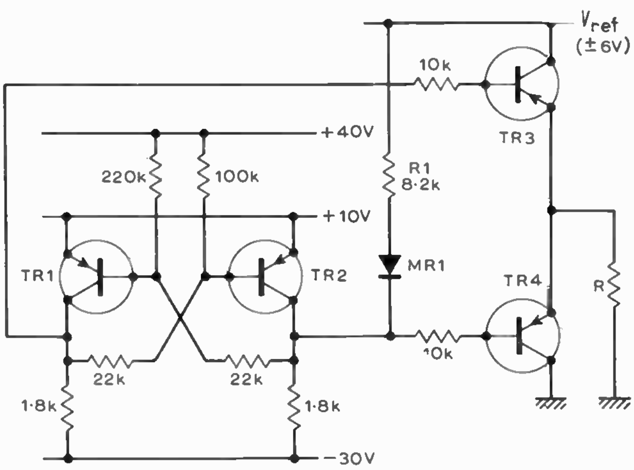


Fig. 2. Bistable and emitter switches in early digital voltmeters.

2.2 *All-solid-state Instruments (The early 'sixties)*

When first introduced, the transistor was not able to replace the valve in all circuit positions. However, the advent of moderately priced silicon transistors with high f_t, low leakage and dissipation limits of above 300 mW transformed the situation and it was now possible to contemplate 'all-solid-state' designs. Strictly speaking the instruments of this generation were not 'all-solid-state' because the display was not solid-state. However, more than adequate reliability was provided by the cold cathode indicator tube and this could be driven directly because transistors such as the 2S745 with a rating of 75 V became available at moderate cost.

Reliability had become something of a problem at this time because developments in connectors and other devices had not kept pace with the improvements caused by the transistor. Figure 3 shows an all-solid-state instrument of the period in which the printed wiring boards are connected to the main frame via edge connectors. However, these edge connectors were used only during assembly and initial test and were soldered rigidly during the remainder of the product life.

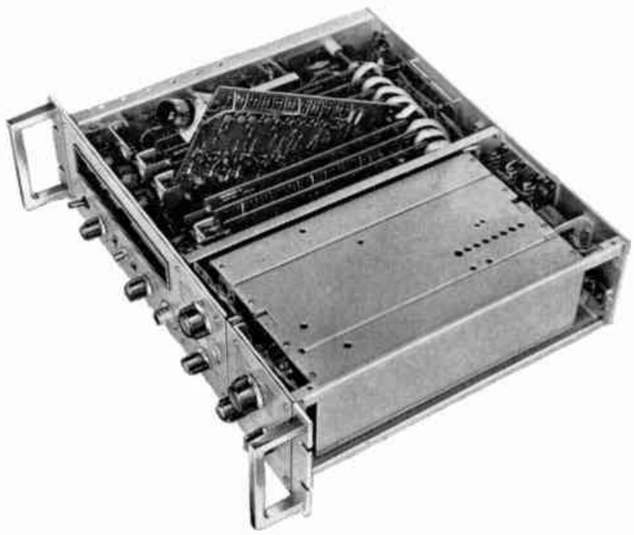


Fig. 3. All-solid-state instruments of the middle 'sixties.

The voltmeter shown in Fig. 3 is also a good example of how the transistor made possible designs which could not be attempted using valves. The principle of operation was to convert the input voltage to a frequency and then accumulate the resulting pulses over a fixed time, thereby finding the average voltage over a defined period (say one mains cycle). To do this reliably required large numbers of low-cost logic elements and this was achieved using the GET882 transistor (later replaced by the D535). With these transistors a bistable element could be built for 50p, and it was feasible to put two or three decades of counting on one printed wiring board.

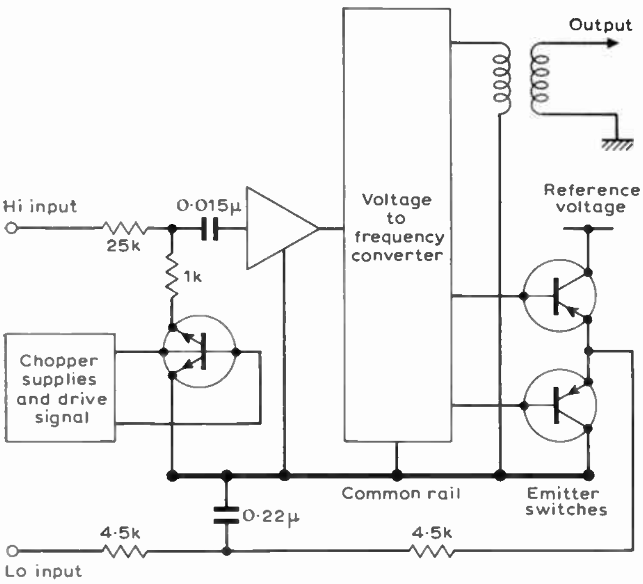


Fig. 4. Input circuit of digital voltmeter which converts voltage to frequency.

The input section was also significantly affected by the use of transistors because a small transformer could be used to transmit the very modest power requirements of the transistors and the input section could readily be made floating and isolated. In fact a common-mode voltage of 450 V caused no error on the 20 mV range.

The high input resistance (greater than 5000 MΩ on the 2 V range) was achieved by overall voltage feedback. Referring to Fig. 4, the amplifier operates on the chopped error voltage and drives a voltage-to-frequency converter. The latter produces pulses of well-defined width which are applied to the emitter switches. Thus the smoothed voltage on the 'Lo' input terminal is a very accurate conversion of the output frequency and the high loop gain ensures both good linearity and a high input resistance. Of particular note in this circuit was the transistor which replaced the mechanical chopper in the earlier designs. This transistor, the 3N72, was remarkable for its time having offset voltages stable to better than 1 μV at moderate cost. These properties were not matched in the planar process for several years.

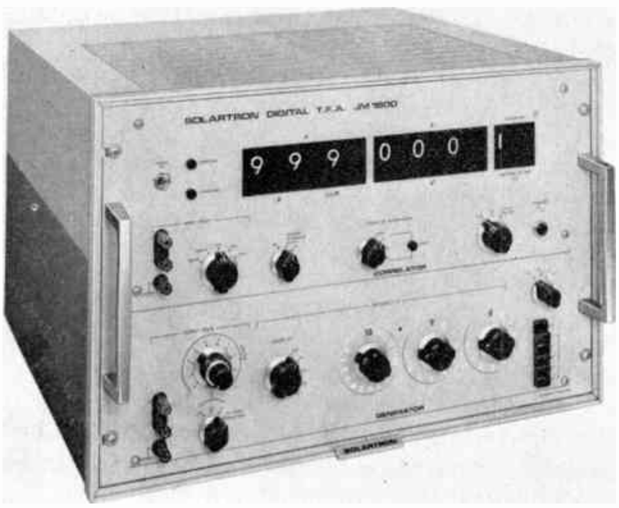


Fig. 5. Digital transfer function analyser containing over 700 transistors and 1900 diodes.

2.3 *Silicon Transistors (The middle 'sixties)*

While silicon transistors had been available for some years and had been invariably used where exacting performance was required, it was not until the middle 'sixties that price reductions permitted their use as general-purpose switching elements. Speed and reliability were two features most evident from this change and designers gladly forgot about thermal instability, but there were some unexpected side-effects. For example, logic circuits now became much more vulnerable to interference transients—particularly on power rails, and terms like 'noise margin' began to assume a practical significance. The problem had not been too evident with earlier transistors, first because they were so slow that there was effectively a 5000 pF decoupling capacitor on each base, and secondly because they had been used in relatively small numbers for simple functions.

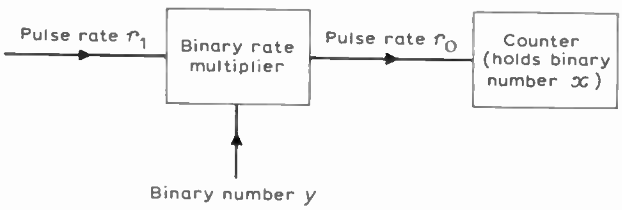


Fig. 6. Digital method of producing $x = \int y$.

Now, however, it was possible to design instruments using large numbers of devices and still achieve acceptable overall reliability and price; the transfer function analyser shown in Fig. 5, for example, has over 700 transistors and 1900 diodes. In this instrument the generator output comes from a digital-to-analogue converter and the correlator input consists of an amplifier and analogue-to-digital converter. All other functions within the instrument are performed digitally. This includes the oscillator and Fig. 6 shows the principle on which the oscillator is based. The binary rate multiplier produces an output pulse rate r_0 which is proportional to the digital number y, r_1 being a fixed rate. The counter accumulates the output from the b.r.m. and thus x, the number in the counter, is proportional to the integral of y. By arranging a second circuit where y and x are interchanged, and given suitable switching, the requirements for simple harmonic motion are satisfied and the x and y registers become the values of sin θ and cos θ. Because the timing is fixed by the clock rate r_1, which can be derived from a quartz crystal, it is possible to generate frequencies like 0·001 Hz and 0·00101 Hz with an accuracy and stability of 0·01 %.

The correlator performs a Fourier integral digitally, the start and stop times for the integral being locked to the generator sinewave. The long-term averaging required for analogue instruments is no longer necessary and answers can be obtained in one cycle of the generator.

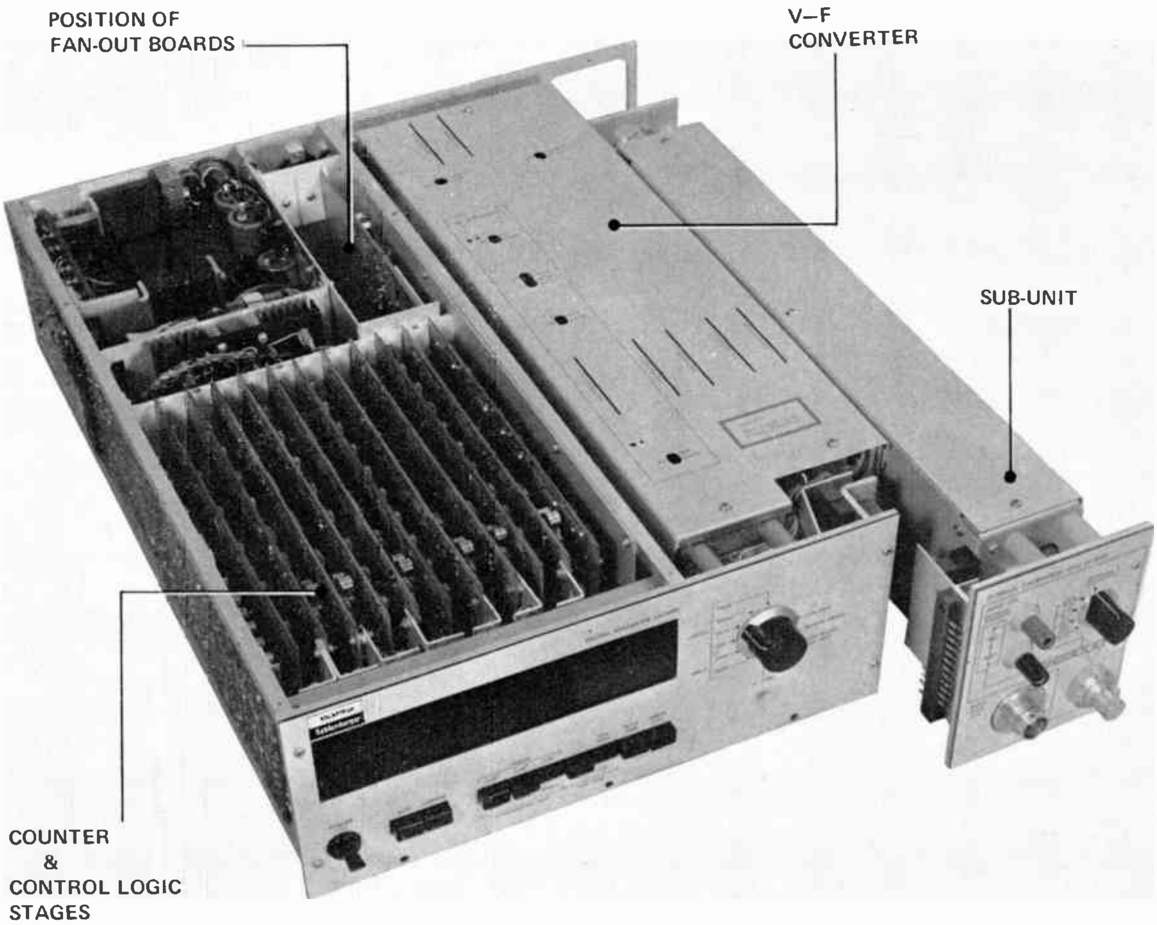


Fig. 7. Integrated circuit digital voltmeter showing good accessibility despite high packing density.

2.4 *Integrated Circuits (The late 'sixties)*

As designers attempted to put more and more digital circuits into instruments, the use of integrated circuits became attractive. Indeed, in terms of reliability their use became almost essential. Prices of these circuits had been falling and as in many other situations instrument designers were able to take advantage of the volume generated in other applications. This accounts for the fact that t.t.l. has been almost universally adopted where many applications could have been satisfied by, for example, r.t.l.

Integrated circuits made possible cheaper instruments and also stimulated the trend towards instruments which are virtually small systems. It is almost impossible to imagine an instrument such as that shown in Fig. 7 without the use of integrated circuits. This voltmeter has an accuracy of $\pm 0{\cdot}001\%$ of range, $\pm 0{\cdot}002\%$ of reading ± 1 μV with a resolution of 0·1 μV. The range is automatically selected and there is an automatic calibration cycle as well as the usual drift correction circuits. For remote operation it is possible to program range, integration period, filter, multiplier and digitize command.

It can be seen, therefore, that increasing degrees of integration first of all made it possible to include more and more automatic functions. However, as costs fell, it became possible to make instruments of the same performance as before but much more cheaply. This particular trend has probably halted and further reductions in cost are likely to come from changing to m.o.s. and related technologies rather than from price reductions on bipolar devices. Already, counters have appeared using m.o.s. circuits but at present the performance is limited to counting rates of 5 MHz without input scaling. It is also to be hoped that these developments will reduce the power consumption because heat has once again become a problem due to the large number of functions which can be provided in one instrument.

3 The Influence of Transistors on Analogue Circuits

3.1 *The Early Years (Up to the end of the 'fifties)*

The difficulties in design associated with analogue instrumentation tend to be characteristic of the product and have less general application than is the case with digital circuits. Nevertheless, the result has been very

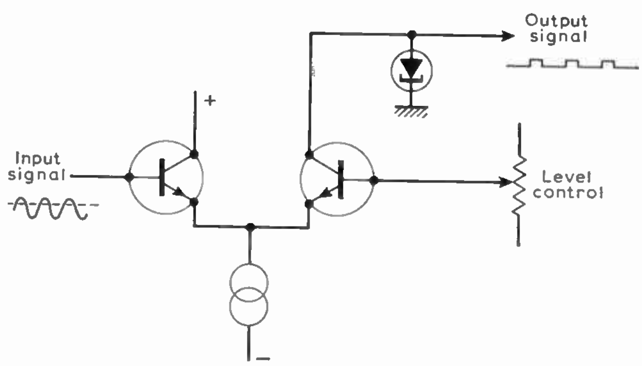


Fig. 8. Tunnel diode level and trigger pulse generation circuit.

similar in most analogue instruments and has meant that transistors were introduced at a relatively late stage after a period of hybrid valve-transistor designs.

For example, the key difficulties in oscilloscope amplifier design are (1) to provide a high input impedance that is well defined, (2) to provide wide bandwidth, and (3) to generate enough voltage swing to drive the tube deflexion plates. The transistor, and particularly the early transistor, was very weak in these three respects.

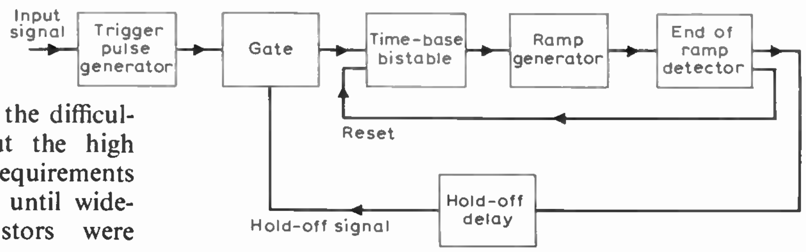


Fig. 9. Block diagram of oscilloscope time-base.

Hybrid valve-transistor designs overcame the difficulties of impedance and voltage swing but the high operating temperatures and bandwidth requirements meant that the transistor had little impact until wide-band ($f_T > 200$ MHz say) silicon transistors were available.

3.2 *The Influence of Tunnel Diodes on Trigger Circuits*

Oscilloscope trigger circuits were more readily adapted to transistor design but the solid-state device which made the most significant improvement was the tunnel diode. Originally developed (unsuccessfully) as a logic element, the tunnel diode can be considered as a two-state current-driven device. It is used in two areas in oscilloscopes where its speed and repeatability are used to the full.

Figure 8 shows the elements of a tunnel diode level and trigger circuit. The input signal is compared with the 'level' generated by a front panel control and the resulting current drives the tunnel diode between its low and high states. Because the tunnel diode is so fast, the output signal is a squared-up version of the input signal even at high frequencies, the tunnel diode transitions taking place at the voltage determined by the level control. In a typical trigger circuit the total swing (say 4·5 mA) required to drive the tunnel diode from its low state to the high state and back again is scaled to correspond to 2 mm on the tube face. As the threshold current (at which a transition takes place) is stable to a small fraction of the 4·5 mA, the trigger point on the tube face is stable to a very small fraction of a millimetre and in practice is limited by the stability of the level control, noise and other interfering effects.

The second position where the tunnel diode has made a marked improvement can be explained by referring to Fig. 9, which shows a block diagram of a simple time-base circuit. The gate is arranged so that a trigger pulse is permitted to drive the time-base bistable only when a ramp has finished and the time-base circuit has settled fully to its quiescent state. Unfortunately there is nothing to prevent a trigger pulse arriving at the gate just as it is being opened by the hold-off signal and therefore the trigger pulse is only partially transmitted. The result is inaccurate timing of the ramp which appears as multiple traces on the tube screen. The most effective way of overcoming this deficiency is to speed up the gating function, and tunnel diodes with transition times in the nanosecond and picosecond range are ideal for the purpose.

3.3 *First All-solid-state Oscilloscopes*

When transistors of reasonable speed and power dissipation did arrive, they were not, at first, used in general-purpose oscilloscopes because they neither provided the ultimate performance at one end of the

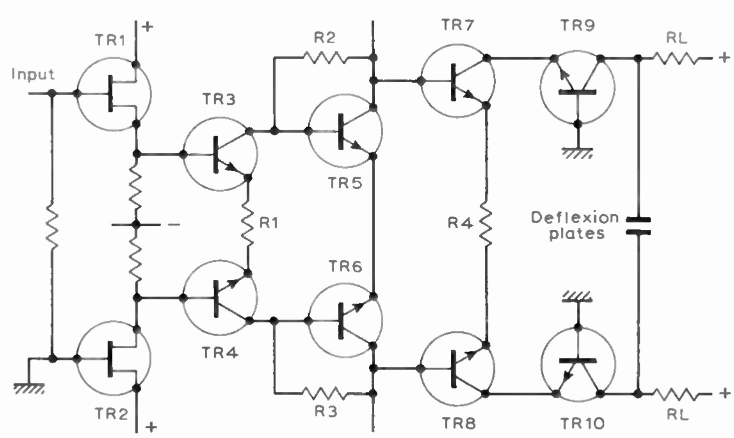


Fig. 10. Outline circuit of an oscilloscope y-amplifier.

scale nor were sufficiently cheap at the other end. The first applications therefore tended to be those where the small size and low power consumption justified a higher selling price, that is, in portable and battery-driven instruments.

Advances in cathode-ray tube technology had been necessary and by the early 'sixties tubes suitable for portable oscilloscopes were available. The field effect transistor (f.e.t.) provided the high impedance required at the input although even at this stage, some designers persisted with Nuvistor input stages.

The amplifier circuit design is of some interest because the same principles could have been used in valve designs with a corresponding improvement in performance. Alternate series and shunt feedback stages were not used in valve designs and in fact many of the early transistor designs were all series.

Figure 10 shows a typical circuit in outline. The f.e.t.s, TR1 and TR2 are connected in common source mode and drive the transistor TR3 and TR4 which have series feedback in their emitters. The next stage transistors (TR5 and TR6) have shunt feedback and they drive another series stage TR7 and TR8. The last stage is a grounded base because of the large voltage swing and high bandwidth required. Alternate series and shunt feedback stages give more bandwidth and better gain stability than either all-series or all-shunt and to a first order the gain depends only on a few resistors (R1, R2, R3, R4 and the plate loads R_L).

3.4 *General-purpose Transistor Oscilloscopes*

Once transistors had been introduced to the portable and battery operated oscilloscopes they provided such an improvement in performance that general-purpose transistor oscilloscopes were not long in following. The price reductions of silicon transistors in the middle 'sixties accelerated this trend as also did the appearance of transistors such as the BFY90 which maintain a very high f_T at low collector voltages.

The need for this requirement can be seen by considering what happens in a simple voltage-driven common-emitter stage which is operated at two sets of d.c. conditions (corresponding to two positions on the tube face). When the operating point of the transistor is changed from one point to the other, the dissipation changes and hence the junction temperature. The base-emitter voltage changes (at about 2 mV per deg C) and thus the operating point makes a further change after a delay depending on the thermal time-constants of the device. In practical circuits this effect (even with negative feedback) is noticeable and is best minimized by choosing the transistor current and collector load so that the transistor operates roughly at constant power. For wideband circuits the load is small and hence the optimum collector-emitter voltage may be as low as 1·5 V. Most transistors suffer a drastic loss of bandwidth at low voltage such as this.

In general terms the introduction of the transistor improved the sensitivity of oscilloscopes from 50 mV/cm to 5 mV/cm and increased the bandwidth by a factor of three.

3.5 *Integrated Circuits in Oscilloscopes*

As is well known, linear integrated circuits have not had the enormous success and growth of digital circuits. In particular, linear integrated circuits have tended to be restricted to operational amplifiers which are unsuitable for oscilloscope *Y* amplifiers. However, operational amplifiers are just what is required for ramp generators and therefore have been widely used in timebase circuits along with standard t.t.l. gates which are suitable for many of the switching and logic functions.

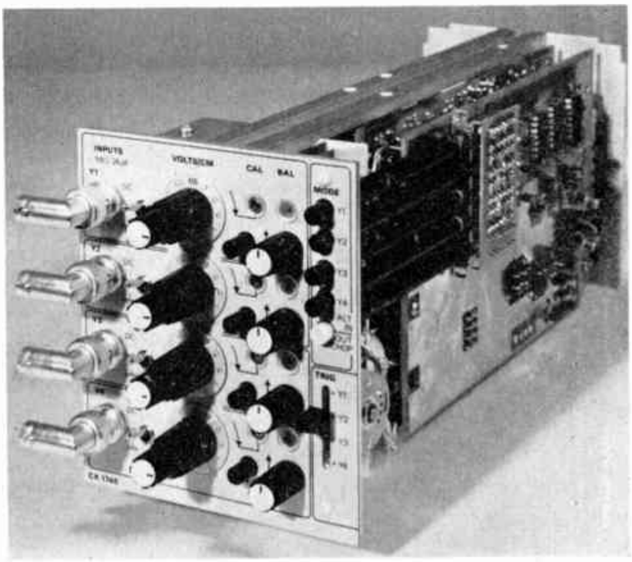


Fig. 11. Plug-in unit using integrated circuits for four-trace oscilloscope.

The combination of properties required for *Y* amplifiers is so specialized that general-purpose amplifiers are unsuitable. These requirements are: direct coupling, wide band width (50 MHz or more), balanced input and output, facilities for electronic switching, addition of bias (shift), gain variation and an auxiliary output for the trigger signal. Custom-designed integrated circuits are a possibility for this requirement and there have been a number of attempts in this direction. One approach has been to use one simple wideband circuit as often as is necessary for the required gain while a second approach has been to put as much as possible into one large integrated circuit per channel. Figure 11 shows what can be done utilizing the resulting packing density to the full. The dimensions of this four-trace unit are only 10·2 cm × 12·6 cm × 25·4 cm (4 × 5 × 10 in) and the module contains the delay line and beam switch.

4 Trends for the Future

While it is difficult to forecast the likely future developments, some trends are already discernable. The first of these is an extension of the systems concept within instrumentation. Low-cost memories, particularly read-only memories, allow the digital instrument to do much that was only possible using digital computers a few years ago. The latter have dropped significantly in price but the technology which has brought this about has also enabled the instrument designer to produce a cheaper, more effective system and one trend appears to be towards more sophisticated instruments with special-purpose computing capability.

A second trend concerns the application of large scale integration techniques to high-volume low-cost products. In this case the initial investment in men and materials is high so that only the companies which already have a large fraction of any one market can contemplate such a proposal. Where the technique is possible the result for the customer is an instrument with better performance at a significantly lower price. Some idea of the evolution of digital voltmeters in this respect can be obtained from Table 1 which shows how the performance and cost have varied over the last 14 years. The effects of inflation are not shown but should be well understood.

Table 1. Evolution of digital voltmeters

Year of introduction	Principal technology	Accuracy	Sensitivity	Price in year of introduction
1959	Ge with relays	0·5%	1 mV	£295
1961	Ge	0·01%	10 μV	£730
1964	Ge and Si	0·05%	10 μV	£425
1968	Si	0·01%	1 μV	£750
1970	Si and i.c.	0·002%	0·1 μV	£1750
1973	m.o.s.	0·03%	10 μV	£195

A third trend appears to be the increasing tendency to use digital circuits in place of analogue circuits, the present range of frequency synthesizers being a typical example. This tendency has been evident for many years but shows no sign of ending although one wonders how justified it is in many applications.

5 Acknowledgments

The author would like to thank all the colleagues who helped in the preparation of this paper by discussing their recollections of the last fifteen years and particularly H. A. Dorey and J. R. Pearce who kindly read the manuscript and provided helpful criticism. The paper is published by permission of the Directors of the Solartron Electronic Group.

Manuscript received by the Institution on 30th October 1972. (Paper No. 1505/IC.71.)

UDC 621.382:378

The influence of semiconductors on the teaching of electronics . . .

A review

K. J. DEAN,
M.Sc., Ph.D., F.Inst.P., C.Eng., F.I.E.E., F.I.E.R.E.*

SUMMARY

Quotations from textbooks in electronics are used to illustrate the impact on education of achievements in semiconductor physics. This work led to the development of transistors and integrated circuits and to the changing fortunes of the microelectronics industry. It has also been responsible for continuous modification of the subject matter of electronics courses. An attempt is made not only to review these changes but to suggest what implications this process may have for the future.

**Principal, South East London Technical College, Lewisham Way, London SE4 1UT.*

1 Education for Electronics before 1948

The story of semiconductor electronics since 1948 is but the more recent part of a much longer tale that stretches back to Faraday's identification of semiconducting properties in silver sulphide in 1833, the Marconi-Kemp experiments of 75 years ago from which telecommunications developed, and the work of Edison and de Forest who laid the foundations of thermionics. Whilst this is not the occasion to review the significance of such work in developing the electronics industry, some account of pre-1948 electronics, particularly with reference to semiconductors, as seen in the textbooks of the day, will set the scene for what is to follow.

Many mature engineers today look back to World War II as their first introduction to electronics. The small radio industry then was based on the domestic market and there was a desperate shortage of engineers. The physics-with-radio degrees of London and other universities and many of the courses specially mounted at technical colleges for newly enlisted men and women were aimed at people entirely new to radio. Courses, often only a few months in length, were planned to satisfy the needs of the armed services more than those of wartime research.

One of the major textbooks of those days, almost the only one readily available, was the two-volume 'Admiralty Handbook of Wireless Telegraphy',[1] published in 1938 and based on an earlier edition revised by Earl Mountbatten. Referring to detectors, it says 'Carborundum and steel is a crystal–metal combination often found in earlier days. A fragment of the crystal is mounted in a small brass cup and the metal contact of cat's whisker was held against it by a spring. The sensitivity of the combination depends very much on the contact pressure and the sensitivity of different "spots" varies. The theory of the action of crystal couples or crystal–metal combinations cannot be said yet to be thoroughly understood'. Indeed it could not! The use of semiconductors like carborundum and galena were only pre-dated as detectors by the coherer. However, with the advent of valves they had been less often used. Few conventional metal rectifiers (e.g. copper oxide) were capable of competition with cat's whisker rectifiers at radio frequencies. The principal exception was the Westector,[2] a miniature stacked-disk copper oxide rectifier which could be used at low radio frequencies.

Wartime research was naturally concerned with immediate solutions and had neither time nor inclination for wide publication. A strong Soviet team which was looking at fundamental semiconductor properties appeared to have been disbanded in 1942,[3,4] whereas US, European and UK activities seemed strongest in 1945–46.[5–8] Most work concentrated on the use of germanium, silicon and boron for detectors, and on a variety of compounds. Doped boron proved unsuitable and maximum operating frequencies with germanium, typically 30 MHz, although suitable for communications were often too low for other purposes. The GEC 'red dot' diode was made from highly-refined silicon with small traces of aluminium and beryllium added. By 1943 silicon with 1 part in 10^5 boron was being used and germanium point-contact diodes were being made with peak inverse voltages of 100 V.

The development of radiolocation, now 'radar', called for higher frequency operation than was common at that time and hence required detectors to operate at these higher frequencies. Valves were, of course, most often used but the transit times of electrons across the valves needed to be short at these high frequencies. This requirement meant that researchers turned once again to the point-contact diode. When the war was over radar and IFF (interrogation, friend or foe) sets were operating at hundreds of megahertz and normal aircraft communication was on bands well in excess of 100 MHz. F. E. Terman,[9] probably the best-known American author of this period, records that both germanium and silicon were in use as detectors and were extensively used in measurement work, particularly at high frequencies. Even so, a good explanation of rectification by semiconductors was not widely available in texts for students' use.

With the war over, returning servicemen who had had experience with radio or radar often regarded electronics as an attractive area in which to make a peace-time career. So technical colleges continued to offer courses in this subject. The 'Royal Signals Handbook of Line Communication'[10] published in 1947 and often used by these students said, 'Considerable research has been carried out to discover where the asymmetrical resistance is introduced. The materials and alloys are all linear resistances . . . the all-important asymmetry of the rectifier is thus assumed to be due to a layer . . . termed the "barrier layer" and although its existence has not yet been definitely proved, it gives a satisfactory explanation of the rectifier action.'

2 Thermionics

Despite any implications to the contrary which this account might have so far given, the emphasis in pre-1948 electronics was on circuit design using thermionic valves. Valves were characterized by high input impedances and voltage amplification. They were relatively fragile, used heated filaments either in evacuated glass envelopes or metal envelopes with a glass–metal seal. They took up much more space than transistors and were then regarded as the least reliable component of an electronic circuit, with the possible exception of electrolytic capacitors. The immediate reaction when a circuit fault occurred was 'Change the valve!' A typical amplifying valve would have a heater dissipating 2 W whilst its anode circuit represented a dissipation of 0·25 W to the high voltage, typically 250 V, supplies.

The transistor was first described in a letter published in July 1948 and with one exception[11] attracted little notice in the press. Philips were said to have produced a working transistor within one week of this notification, G.E.C. and Sony started research projects whilst most other companies waited for the Bell Symposium in 1952 before making much progress. The vital work of J. Bardeen and W. H. Brattain under the direction of W. Shockley at Bell Telephone Laboratories which led to the discovery of the transistor has been described elsewhere and can be seen in better perspective now. (Most of the early papers on transistors were reprinted in 1969 with additional comments by E. J. M. Kendall.[12]) An extensive bibliography of papers during this period was also published by Pye Ltd in 1954.[13]

By 1954 most of the companies now in semiconductors had 'set up shop' and there were tangible indications of progress with transistors. For this reason it is particularly interesting to read a small book published by Newnes in 1954 over the names of M. G. Say and E. Molloy[14] which claimed to be written by specialist contributors for the professional engineer who needed the latest information. With reference to the view that crystal triodes and thermionic valves were complementary it added 'It is too early to foretell the extent of the use of the crystal valve and how far it will affect the use of the thermionic valve. The general opinion seems at present to be that it will displace its established competitor only in some special cases.' Referring to problems of manufacture delaying the move into semiconductors it states 'Even then the general introduction of transistors will be gradual, determined by cost and by the continued success in overcoming technical problems.' One can only add—it's later than you think!

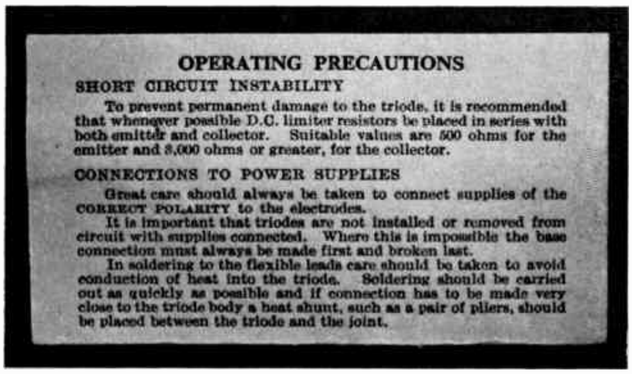
One would have thought, with the wisdom of hindsight, that these small rugged transistors would have been seen to be preferable to valves, even though they were expensive and limited to operating at relatively low frequency and power. But there were no heaters and a typical circuit dissipation might be under 10 mW—a factor of improvement of 250 on valves. There were also signs of better things—an article in 1950 described the use of point-contact transistors at 23 MHz[15] and one two years later described a 500 mW push-pull power amplifier.[16]

In fact it was nearly eight years before transistors were 'discovered' by education and courses in transistor electronics became popular in the technical colleges, and even then they were often tacked on to *ad-hoc* thermionics courses and employed the established 'known-to-the-unknown' teaching principle of valve analogy. Even in 1973, 25 years on, there are probably isolated cases where this practice of thermionics first has not been entirely abandoned. New ideas have always been resisted in education. However it is perhaps as well that, within reason, one should not be too keen to bow to every wind of change.

3 The Institution Looks at Transistors

The Institution, formerly the British Institution of Radio Engineers, has always regarded education as an important function. Past copies of the Graduateship examination (now replaced by CEI part 2) were even bound in with the *Journal* in 1952 when it published a paper 'Crystal triodes' by E. G. James and G. M. Wells.[17] The authors, from the GEC Research Laboratories at Wembley, first presented their paper at the South Midlands Centre in April 1951. They describe what must have been one of the first UK-made point-contact transistors (Fig. 1). However they sit on the fence, as well they might so early on, when they predict 'That the crystal triode . . .

will find many applications where it will perform better than the thermionic triode but there are others where it is never likely to compete.' Unfortunately, the subsequent discussion at the meeting did not elicit details.


OPERATING PRECAUTIONS

SHORT CIRCUIT INSTABILITY

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. Suitable values are 500 ohms for the emitter and 8,000 ohms or greater, for the collector.

CONNECTIONS TO POWER SUPPLIES

Great care should always be taken to connect supplies of the CORRECT POLARITY to the electrodes.

It is important that triodes are not installed or removed from circuit with supplies connected. Where this is impossible the base connection must always be made first and broken last.

In soldering to the flexible leads care should be taken to avoid conduction of heat into the triode. Soldering should be carried out as quickly as possible and if connection has to be made very close to the triode body a heat shunt, such as a pair of pliers, should be placed between the triode and the joint.

(*photo: J. Marrett*)

Fig. 1. The GET 1 was one of the first point-contact crystal triodes to be manufactured in Britain (about 1953). The maximum collector dissipation was rated at 100 mW with a maximum collector current of 15 mA and a current gain greater than 2. The maximum ambient operating temperature was 35°C. In days when most engineers were unfamiliar with semiconductors, the operating precautions printed on the reverse side of the packet in which the transistor was sold, were essential.

4 The Inertia of Education

The concept of inertia in education is a highly emotive one, and one thinks of lecturers using the same set of lecture notes 30 years on! But in fact the changes in electronics have required lecturers and writers alike to rethink radically their approach to their subject and to learn new techniques: a traumatic experience which a few never survived!

The inertia of education shows itself in two ways: an unwillingness to accept new ideas and even when an idea is accepted, a slowness in realizing its implications. There are, of course, some outstanding exceptions to this, both people and topics, but course tutors and textbooks writers were generally not sufficiently farsighted in interpreting the innovation and exploitation of semiconductor devices. P. Parker's 'Electronics',[18] a standard work published in 1950, did not even index the term 'transistor' in the 1963 reprint and appeared virtually to disregard solid-state electronics. F. E. Terman's 'Electronic and Radio Engineering'[19] in the 1955 edition devotes only one of over twenty chapters to transistors, a mere 70 pages out of 1078 pages in all to point-contact and junction devices, stressing particularly the high α, in excess of unity, of point-contact transistors. E. T. A. Rapson's 'Experimental Radio Engineering'[20] had a chapter added on experiments with transistors (11 out of a total of 104) in the 5th edition in 1964. M. G. Scroggie is another writer whose 'Radio and Electronic Laboratory Handbook'[21, 22] was so well known. First published in 1938, the 6th edition in 1954 had no reference to transistors whilst the 7th edition in 1961 compared valves and transistors but advised that 'Laboratory oscillators which are required to perform within close limits can more easily be designed around valves.' With reference to the small size and low power of transistors, it added 'Where size and portability are important their disadvantages may well be tolerated or overcome.' It was only in the 8th edition in 1971 that Scroggie could comment 'By now transistors and many other semiconductor devices are the norm. The use of valves . . . needs to be justified by special circumstances.' Nevertheless even this edition[23] has a surprisingly large number of diagrams showing valves.

By 1961 there were signs that even textbook writers were beginning to be aware of the implications of semiconductors. D. K. McCleery[24] in 'Radio Today' in comparing transistors and valves states 'They are at present rather more expensive than ordinary valves and there is the frequency limitation . . . for junction transistors. Transistors are also less uniform in their characteristics than vacuum valves, although the crystal diode can be used up to the highest frequencies.' All of this was fair comment for say, 1958, but the changing climate of opinion was echoed by the remark, 'There are some who forecast confidently that by the middle 1960's there will be no valves at all in receivers.'

The watershed was passed by the time the writer's own 'Laboratory Manual of Electronics'[25] went to the publishers in 1965 for 41 out of the 66 experiments in it used solid-state devices and the list of experiments commenced with 'Parameters of transistors and valves'. Three years earlier the order might well have been 'Valves and transistors'.

One must always remember however that from the time an author sets out to write a textbook to the time it appears in print two years or longer may have elapsed. Nevertheless there is less excuse for inertia in new editions of established texts; and even in a first edition really important additions can be put in at galley proof stage.

If there is one place where inertia in education is brought to light more than in any other, it is in those textbooks which consist of past questions from examinations by universities, professional bodies such as IEE, IERE, and IoP, and the internal examinations of colleges. C. S. Henson, perhaps the best known of such compilers, first published his 'Solutions of Problems in

Telecommunications and Electronics' in 1956, the 1962 edition[26] containing 11 out of 330 pages on transistors, and the 3rd (1968) edition[27] having 18 pages out of 406, or 3·3% and 4·4% respectively. This gross imbalance becomes 'fed back' to the next generation of students. In periods of drastic change the content of these examinations changes also and the value of such books must be open to severe doubt.

5 The Changing Technology

Amongst the first changes brought to education by the new technology was the introduction of semiconductor physics into applied physics courses for engineers as well as *ad-hoc* transistor courses, usually on an evening or part-time basis. Professional engineers flocked to them in the mid-1950s often at their own expense, sometimes going from one college to another to extend their knowledge. Higher National Certificate and Diploma courses began to include transistor electronics in the final year, based on the subject matter in the successful evening courses: degree courses, for example the London external degree, were slower to respond. Understandably it has never been easy to change the syllabus quickly of an examination for which students are prepared over a number of years by many widely dispersed colleges. A lasting change has been that materials science is now an integral part of engineering courses.

6 Circuit Theory

The more obvious innovations in subject matter in the new technologist and technician courses were the use of a variety of 2-port equivalent networks instead of the Thevenin's and Norton's circuits used to represent the behaviour of valves, often whole chapters being devoted to this topic alone.[28] K. W. Cattermole[29] writing in 1959 in one of the first books which brought a more modern approach to transistors, said 'The transistor unlike the pentode valve is far from being a one-way amplifier; its internal feedback is rarely negligible and sometimes dominant.' As high-frequency operation became possible the limitations of the equivalent circuits became more obvious and at one time the number of parameters, and the manufacturers' use of them to show their products in good light, was quite bewildering. This applied particularly to the use of f_α, f_1, f_T etc. Definitions of high-frequency gain sometimes hid what a graph would have revealed. No wonder E. H. Cooke-Yarborough[30] wrote in 1957, 'The development of electronic circuits is a bad scientific discipline because to experiment with a circuit is often easier and quicker than to calculate its properties in detail.'

One advantage of the 'scaling down' of resistor values and supply voltages from those used with valve circuits and the development of both n-p-n and p-n-p devices was that directly-coupled amplifiers were now a reality. The use of current amplifiers rather than thermionic voltage amplifiers meant that amplifier design was not in fact merely the scaled-down version of the valve circuits. Feedback theory was better understood and taught, largely due to a paper by E. M. Cherry[31] which first appeared in the *Proceedings of the Institution of Radio Engineers, Australia* and for which it won their Hayes Memorial Medal in 1962 by the recommendation of this Institution, and hence appeared in *The Radio and Electronic Engineer* in 1963.[32] By the following year its substance was being written into textbooks,[33] so laying the groundwork for the emergence of the operational amplifier as a building unit of analogue systems.

Other changes in circuit design have followed from the development of integrated electronics. Absolute component values were not easy to obtain with any degree of certainty; it was much easier to obtain the relative values of circuit elements. Hence design techniques were developed by manufacturers which relied upon maintaining relative values rather than absolute ones. In the colleges increasing emphasis has been placed on 'worst case' and statistical design methods. However, the wider issue of the relevance of circuit design should be questioned in the context of an electronics industry where few people are employed on this work, if only to clear it of the charge of irrelevance.

Now that we are in the era of large scale integration and circuit design is becoming increasingly complex, people are turning to computer-aided design of circuits and chip layout—topics which would never have been thought worthwhile with valves or even with discrete transistors.

7 Cross-fertilization of Ideas

One must not under-estimate the part played in education by those being educated and by the training officers who sent them to the colleges. Even if textbook writers and educationalists were slow to develop the new technology some sectors of industry were not. And it was from industry that HNC and HND students came into the colleges with their new ideas. Although in 1955 T. R. Scott[34] could write 'Hesitancy in use is leading to hesitancy in manufacturing programmes' when referring to the 'condition of incipient stalemate' which he said existed, ten years later it was all vastly different. The cut and thrust of industry was making the price of competing in semiconductor manufacture ever higher and the price of the transistors they made ever lower. If the same state of affairs had existed with car manufacturers, a 'Mini' would today be selling for the price of a second-hand bicycle!

All of this was brought about by a mixture of careful research and happy chance and by the support of funds deriving from military confrontation and the space race between USA and USSR, so graphically portrayed in the OECD report 'Gaps in Technology—Electronic Components'.[35] It is fascinating to read the saga of companies like Philco where mistakes in technology and management tumbled them from leading positions in world markets and in some cases to eventual extinction; but the cross-fertilization of ideas in educational circles resulted in lecturers keeping in touch with industry and bringing examples of technological developments in electronics into the classroom.

8 Integrated Circuits

Transistors were attractive because of their small size, but at first only those with defence contracts could

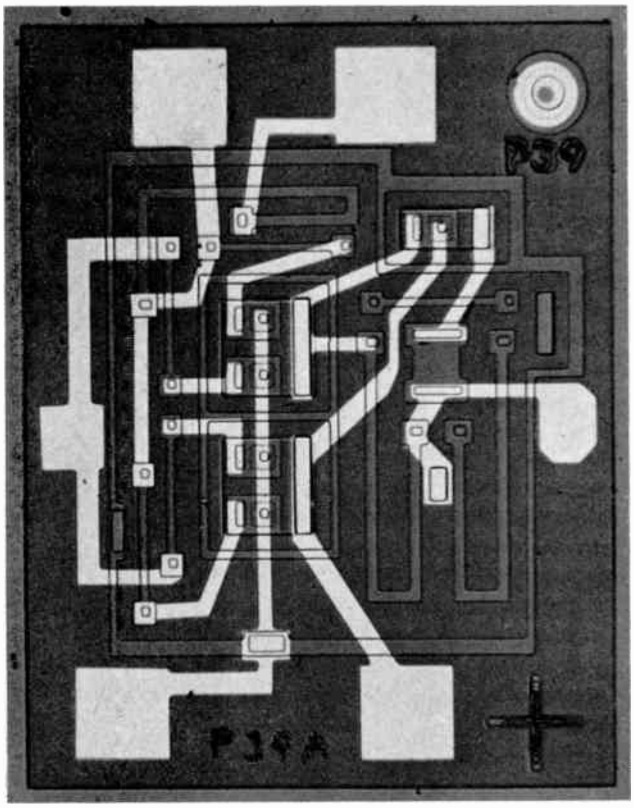


Fig. 2. Early monolithic planar integrated circuits available in 1962 consisted of about 10–15 active and passive components. The circuit shown here was 1·0 × 1·2 mm in size and comprised 5 transistors and 7 resistors. The first digital circuits usually employed resistor-transistor logic, and the first linear circuits were comparator amplifiers.

afford them. Greater skill with technology resulted in falling prices, more precisely defined characteristics and greater reliability. Today no undergraduate course in electronics would be complete without some treatment of reliability and its evaluation. The subject has also become an important part of City and Guilds' Technician Courses. J. E. Shwop's 'Semiconductor Reliability'[36] opens with the words 'The importance of semiconductor reliability would be difficult to over-estimate.' It is this feature of high reliability perhaps more than that of merely putting together large systems and so overcoming the tyranny of numbers that is so characteristic of modern electronics and distinguishes it from the era of valve circuitry.

Size, price and reliability were all so attractive to users that the success of transistors was at last assured. The development of mesa transistors and then planar technology in 1958–59 made possible the integrated circuits which had been proposed by G. W. A. Dummer[37] in 1952, and the 'solid circuits' discussed at the symposium on electronic components held at the Radar Research Establishment at Malvern in 1957.[38] There were two effects which integrated circuits had on education: they further emphasized the development in industry of digital techniques, which had been going on since World War II and which were already important with transistors, so leading to closer attention by colleges to digital electronics, and they showed that the concept of a component had altered in complexity from a resistor, capacitor or transistor to a circuit such as a logic gate or an operational amplifier, so encouraging the modular approach in electronics (Fig. 2). This is not to say that

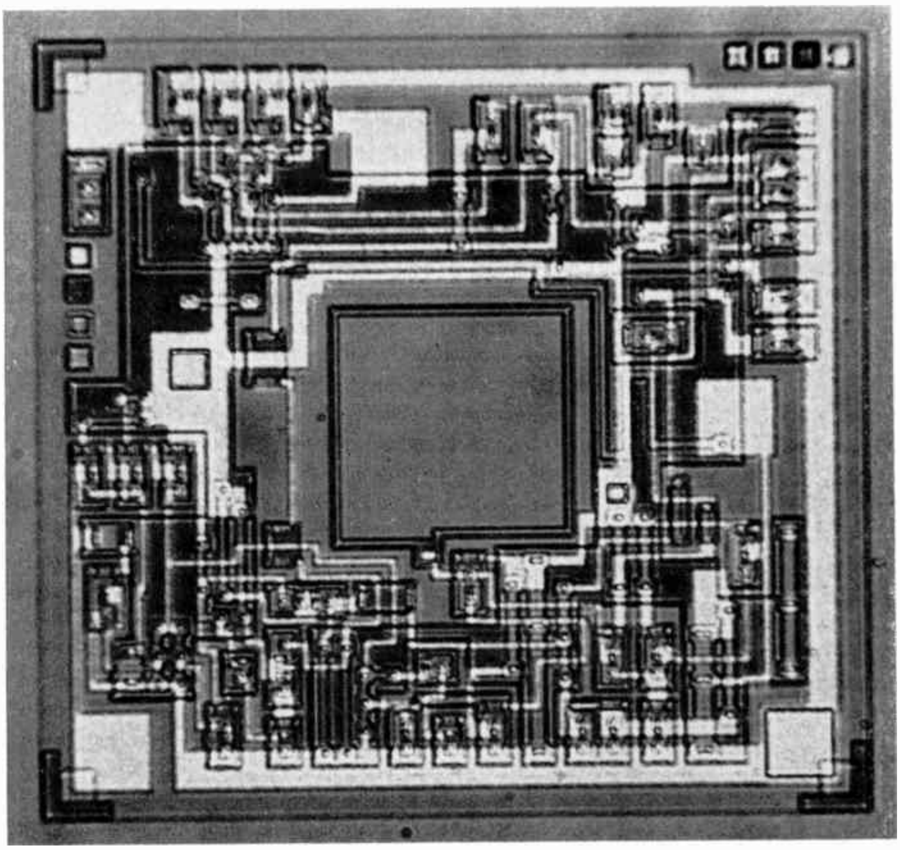
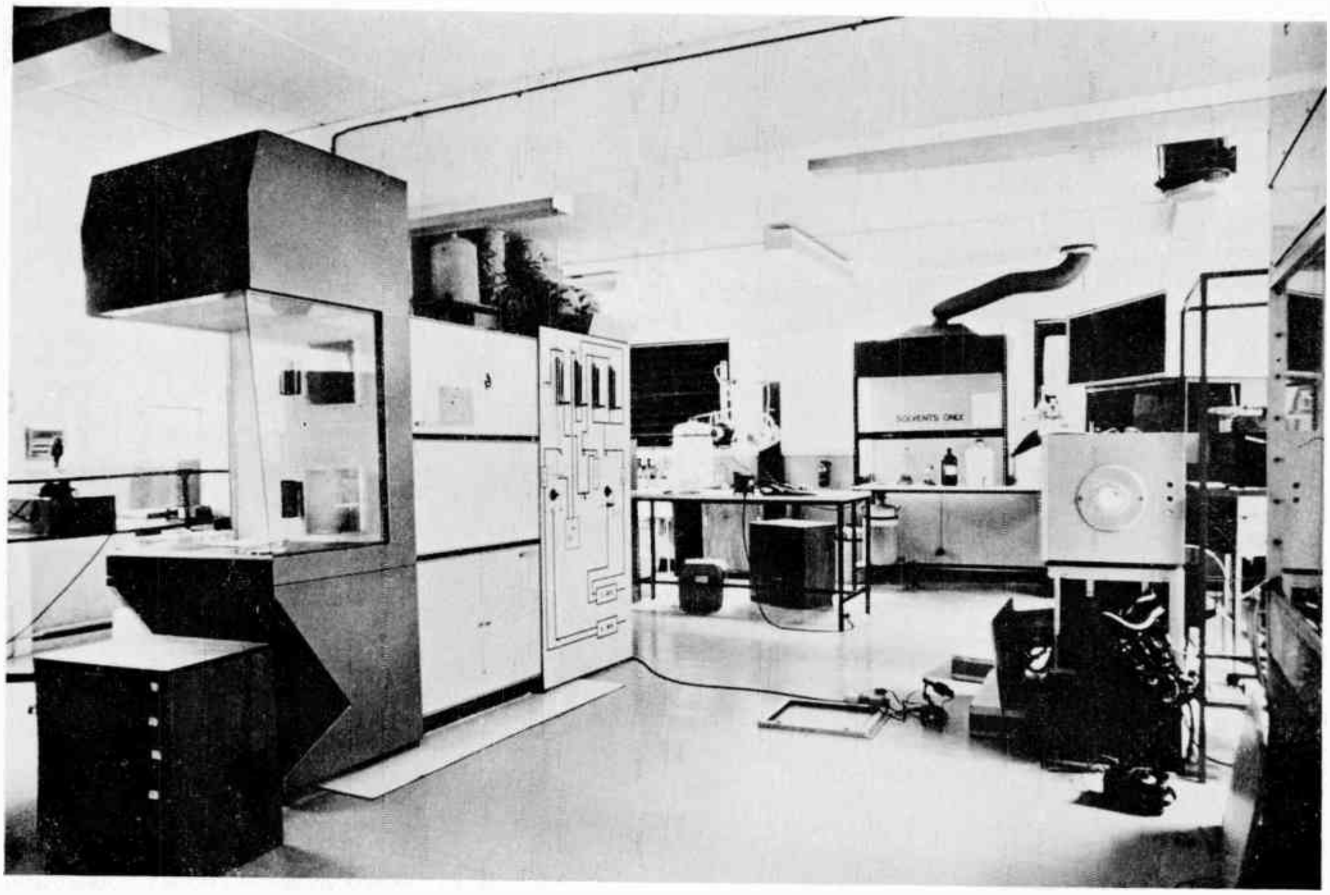


Fig. 3. By 1972 a 1·0 mm square chip like the bipolar photoswitch shown here (by courtesy of Ferranti Ltd) contained about 50 active elements alone. The photocell in the centre takes up over 10% of the surface area of the chip. The greater density of component packing (see Fig. 2) makes it possible to include the photocell, comparator, oscillator, and output flip-flop on the chip, leaving correspondingly less for circuit designers to do externally.

(photo: E. Forbes)



(photo by courtesy of Edinburgh University)

Fig. 4. The furnace room of the microelectronics unit in the School of Engineering Science at the University of Edinburgh.

interest in semiconductor physics, for the general student[39] or the specialist was any less. A. A. Shepherd's excellent 'Introduction to the Theory and Practice of Semiconductors'[40] in 1958, had its counterpart ten years later in H. R. Camenzind's 'Circuit Design for Integrated Electronics'.[41] Nevertheless as usual educators were slow to react: at a widely advertised UK electronics conference for lecturers in higher education in the London area in 1966 to discuss integrated electronics only 20 people were sufficiently interested to attend,[42] although industrial conferences operated with a much higher fee, like those on 'Microelectronics in Equipment',[43] held in 1966 and 1967 attracted hundreds of delegates. A conference organized by SGS-Fairchild in 1967 however, 'Education for the Microcircuit Era', intended primarily for university and college lecturers, was more successful and over 200 attended.

The combination of digital electronics and integrated circuits has for good or ill tied the microelectronics industry to the fortunes of computer manufacture, for the modular concept was appreciated from the start in computer design: it has also led to new circuits. The 'special circuits' of World War II (e.g. the *kipp* relay and the flip-flop) have been developed first into modular discrete transistor logic circuits such as Norbits*, Combi-elements* and Mini-log†, then into integrated circuits—master-slave JK flip flops and similar circuits—not because they are more versatile, though they may be that, but because these specific functions could be sold in large numbers to computer manufacturers. Computers could not have developed as they have without transistors and integrated circuits. Thus now we teach set theory as part of school mathematics, with Boolean algebra, Karnaugh maps, the design of registers and counters, race hazard theory and much else of the same sort in our colleges.

* Mullard trade name. † GEC-Elliott trade name.

9 Practical Work

Some universities, notably Edinburgh and Southampton, and also the Enfield site of the Middlesex Polytechnic developed microelectronics suites for the small-scale manufacturer of semiconductor devices (Figs 4, 5, 6). This is a facility which British education should not be without. However the rapid progress in technology has made it difficult to keep up with industry. Indeed the protagonists of this work would agree that the training they give must be supplemented by industrial experience, if for no other reason than that they are training for that part of industry. However it is sadly true that the number of places where this skill is useful is much less today than when they were set up. Plants at Glenrothes and Witham for example are now closed and it has been said that one manufacturing site could probably satisfy the present needs of British industry. Whether this will be

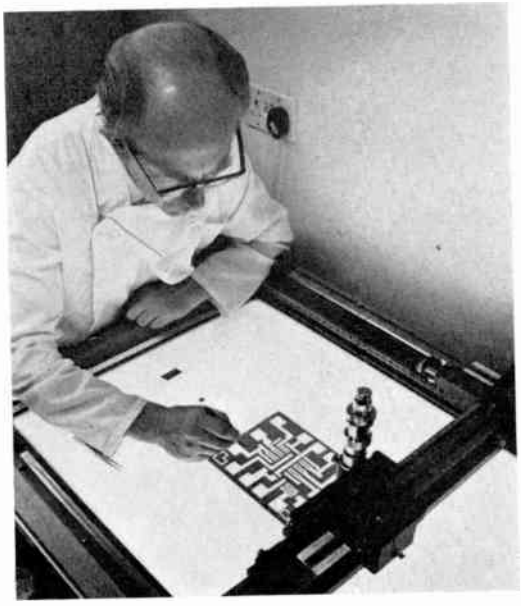


Fig. 5. Cutting artwork for an integrated m.o.s. circuit at the Microelectronics Centre of the Enfield College of Technology, now Middlesex Polytechnic.

so in future will depend on the extent to which diversification away from computing is successful and on our ability to compete in European and world-wide markets. The eventual change which integrated electronics has brought about is that a much smaller number of very good circuit designers are needed now than formerly. The much larger problem is the training of all those other engineers who are not involved in processing semiconductors.

On the applications side of college practical work there has been a tradition of 'bread-board' experiments on which simple circuits have been laid out so that measurements can be made of circuit performance. The effect of integrated circuits has been to exchange these simple bread-boards for more sophisticated ones, either commercially produced[44] or 'home made' based on products like multi-layer Selectroboard[45] (Fig. 7). Initially this had its chief impact in digital electronics. Practical work for logical design is now almost exclusively the art of interconnexion of integrated circuit packages without degrading their performance. Thus W. A. Lo[46] could write in 1967 'With the advent of solid circuit technology the device-circuit boundary is rapidly vanishing. This is most evident in integrated electronics. . . .' As more m.s.i. chips are used even this may prove to be too limited a view, the relevant boundary now in question being that between circuit and sub-system.

Work with integrated linear circuits was slower to develop. There was no ready-made customer capable of absorbing large numbers of chips for linear amplifiers as the computer market absorbs digital chips. Nevertheless here too the parameters of an integrated operational amplifier have become of more importance than those of the components of which they consist. Thus W. Gosling[47] in a book dealing almost exclusively with integrated amplifier circuits wrote 'It is doubtful whether the user of an integrated circuit need necessarily concern himself at all with the devices of which it is composed.'

10 Education Today

There were a number of papers presented at the Institution's Cambridge Conference in 1968 'Electronics in the 1970s' which were relevant to the effects of semiconductors on education. D. F. Dunster and T. Wilmore[48] describing a microelectronics suite at the N. E. London Polytechnic conceded 'There is no case for teaching electrical engineers to be device technologists . . . the technology is too difficult.' The problem which they went on to describe was that, 'It is impossible to do electronics laboratory work *on* integrated circuits, as distinct from systems work with i.c.s.' This is one part of the

Fig. 6. Loading thick film circuits into a furnace at the Enfield Microelectronics Centre. Circuits like this might carry silicon integrated circuits, connected as flip chips or by beam leads, to form hybrid systems.

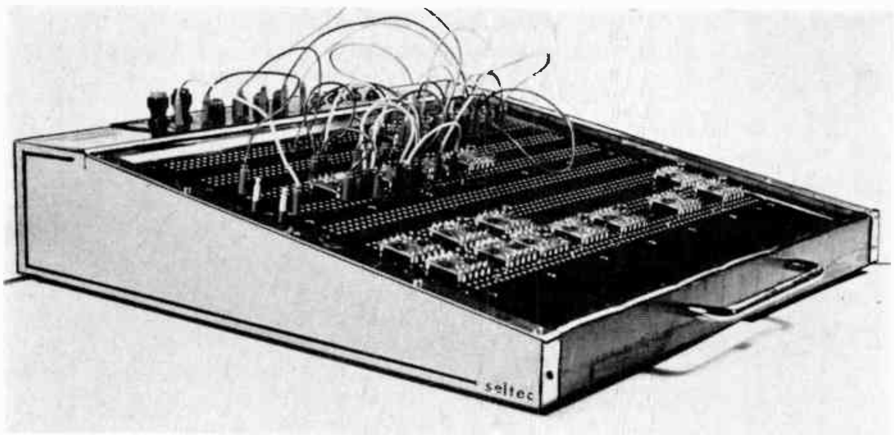


Fig. 7. With this college-made 'bread-board' an integrated circuit can be mounted on a carriage and plugged into the 2-layer matrix board. The longer pins on the carriage connect to the power supply layer of the matrix and the shorter pins can be used to interconnect signal leads. In this way small-signal systems can be patched together.

problem for practical work in the future. The rest of the problem can be illustrated by a paper, also at the Cambridge Conference, given by B. H. Venning.[49] In it he commented '. . . already it is clear that many of our electronics engineers at present in college will become systems engineers, probably none being called upon to design individual circuits but working with linear or digital integrated modules.' Why then, we might ask, is it the case that in many of our undergraduate and technician courses the technique which is being most strongly pressed is that of circuit design, if this is precisely the area which will not be needed? A recent editorial in *Electronic Engineering* commenting on the record of 50 years development in electronics posed the question 'Why should we expect electronics to be there 50 years hence?' The way things are going in practical engineering terms we might all echo, 'Why indeed?'

11 What of the Future?

A manpower study[50] carried out in 1967 referring to the electronics industry reported 'The rate of technological innovation is outstanding even for a science-based industry . . . microcircuits are seen as potentially by far the most important technical innovation.' We have seen this borne out, with developments since then leading to reduction in the area occupied by devices on a chip and increasing yields so leading to more complex systems on the chip. Most people would agree that the one thing that is important to teach is fundamentals. But what is fundamental? Is circuit design as worthwhile as system design? New techniques, like collector diffusion isolation are making digital and linear circuits compatible on one chip so that whole systems can be integrated. A study of small computers[51] carried out three years ago stated 'In the ten years to 1980 the dominant effect on computer internal hardware, architecture, production and manufacturing will be the continuing development of integrated electronics.'

The modern hand-held calculator market is also based on single-chip systems philosophy. Other areas for exploitation are the white goods market (e.g. washing machine controllers), the automobile trade (at £30 potential for electronics per car this is an important area), and domestic entertainment. There will soon be so little electronic engineering involved that the engineering of interconnexions will be all-important. These are ever-opening vistas for industry, but the role to be filled by education, particularly at undergraduate level is perhaps not nearly as clear as it was in 1948. Then it was only a matter of keeping up. Today with complete systems being available on a single chip, and once engineered prices falling to a tiny fraction of their original market value, repair being both impracticable and uneconomic, there may be almost nothing to keep up with at all!

12 References

1. 'Admiralty Handbook of Wireless Telegraphy' (B.R. 230, H.M.S.O., 1938). (Earlier editions were dated 1925 and 1931.)
2. Sowerby, A. L. M., 'Foundations of Wireless', 3rd Edn. (Iliffe, London, 1942).
3. Davidov, B. I., 'Transitional resistances of semiconductors', *J. Phys., USSR*, **4**, pp. 335–9, 1941.
4. Davidov, B. I. and Gurevitch, B. Kh., 'Voltage fluctuations in semiconductors', *J. Tech. Phys.*, **12**, pp. 31–5, 1942.
5. Seitz, F., 'The basic principles of semiconductors', *J. Appl. Phys.*, **16**, pp. 553–63, October 1945.
6. Busch, G., 'Electrical conductivity of silicon carbide', *Helv. Phys. Acta*, pp. 167–88, May 1946.
7. 'Germanium alloys' (Abstracts from Americ. Phys. Soc. papers), *Physical Review*, March 1946.
8. For a wide range of papers in the field see ref. 13.
9. Terman, F. E., 'Radio Engineering', 3rd Ed. (McGraw-Hill, New York, 1947).
10. 'Royal Signals Handbook of Line Communication' (HMSO, 1947).
11. 'Amplifying crystal,' *Wireless World*, **54**, p. 358, October 1948.
†12. Kendall, J. M., 'Transistors' (Pergamon, Oxford, 1969).
†13. Meyrick, N. L. and Roman, G., 'Ten Years of Semiconducting Materials and Transistors' (Pye, Cambridge, 1954).
14. Say, M. G., 'Crystal Rectifiers and Transistors' (Newnes, London, 1954).
15. Brown, C. B., 'High frequency operation of transistors', *Electronics*, **23**, pp. 81–3, July 1950.
16. Shea, R. F., 'Transistor power amplifiers', *Electronics*, **25**, pp. 106–8, September 1952.
17. James, E. G. and Wells, G. M., 'Crystal triodes', *J. Brit. Instn Radio Engrs*, **12**, pp. 285–94, May 1952.
18. Parker, P., 'Electronics', 7th printing (Arnold, London, 1963).
19. Terman, F. E., 'Electronic and Radio Engineering', 4th Edn. (McGraw-Hill, New York, 1955).
20. Rapson, E. T. A., 'Experimental Radio Engineering', 5th Edn. (Pitman, London, 1964).

† These references have extensive bibliographies.

21. Scroggie, M. G., 'Radio Laboratory Handbook', 6th Edn. (Iliffe, London, 1954).
22. Scroggie, M. G., 'Radio and Electronic Laboratory Handbook', 7th Edn. (Iliffe, London, 1961).
23. Scroggie, M. G., 'Radio and Electronic Laboratory Handbook', 8th Edn. (Iliffe, London, 1971).
24. McCleery, D. K., 'Radio Today' (Oxford University Press, Oxford, 1961).
25. Dean, K. J., 'A Laboratory Manual of Electronics' (Blackie, London, 1966).
26. Henson, C. S., 'Solutions of Problems in Telecommunications and Electronics', 2nd Edn. (Pitman, London, 1962).
27. Henson, C. S., 'Solution of Problems in Electronics and Telecommunication', 3rd Edn. (Pitman, London, 1968).
28. Abrahams, J. R. and Pridham, G. D., 'Semiconductor Circuits', Chapter 4 (Pergamon, Oxford, 1966).
†29. Cattermole, K. W., 'Transistor Circuits' (Heywood, London, 1959).
30. Cooke-Yarborough, E. H., 'An Introduction to Transistor Circuits' (Oliver and Boyd, Edinburgh, 1957).
31. Cherry, E. M., 'An engineering approach to the design of transistor feedback amplifiers', *Proc. Instn Radio Engrs Aust.*, **22**, p. 303, May 1961.
32. Cherry, E. M., 'An engineering approach to the design of transistor feedback amplifiers', *The Radio and Electronic Engineer*, **25**, pp. 127–44, February 1963.
33. Dean, K. J., 'Transistors: Theory and Circuitry' (McGraw-Hill, New York, 1964).
†34. Scott, T. R., 'Transistors and Other Crystal Valves' (MacDonald and Evans, London, 1955).
35. 'Gaps in Technology—Electronic Components', OECD, Paris, 1968.
36. Shwop, J. E., 'Semiconductor Reliability' (Chapman and Hall, London, 1961).

† These references have extensive bibliographies.

37. See Ref. No. 35, page 59.
38. 'International Symposium on Electronic Components', RRE, Malvern, 1957. For report see (Ivall, T. E.), 'Solid circuits: glimpses into the future at Malvern components symposium', *Wireless World*, **63**, pp. 516-7, November 1957).
39. Seymour, J., 'Physical Electronics' (Pitman, London, 1972).
40. Shepherd, A. A., 'An Introduction to the Theory and Practice of Semiconductors' (Constable, London, 1958).
41. Camenzind, H. R., 'Circuit Design for Integrated Electronics' (Addison-Wesley, Reading, Mass., 1968).
42. Dean, K. J., 'Integrated electronics—its effect on education', *IEE News*, August 1966.
43. Conferences on 'Microelectronics in Equipment', London, November 1966 and November 1967 sponsored by *Electronic Equipment News*.
44. Stearman, G. H., 'Equipment for learning logic', *Control*, **8**, p. 506, October 1964.
45. Dean, K. J., 'Integrated electronics—its place in education', *Electronic Engineering*, **38**, pp. 458–60, July 1966.
46. Lo, A. W., 'Introduction to Digital Electronics' (Addison Wesley, Reading, Mass., 1967).
47. Gosling, W., 'An Introduction to Microelectronic Systems' (McGraw-Hill, New York, 1968).
48. Dunster, D. F. and Wilmore, T., 'The purpose of teaching microelectronics', Convention on 'Electronics in the 1970s', pp. 15–28, IERE Conference Proceedings No. 12, Cambridge, 1968.
49. Venning, B. H., 'The influence of microelectronics developments in education in the 1970s'. *Ibid*, p. 103.
50. 'Manpower Studies, No. 5—Electronics', HMSO 1967.
51. 'Study of Small Computers', Department of Trade and Industry, November 1970.

Manuscript first received by the Institution on 13th October 1972 and in final form on 1st November 1972. (Paper No. 1506/Ed14.)

UDC 621.382.008.05

Dominant trends affecting the future structure of the semiconductor industry

I. M. MACKINTOSH,
Ph.D., C.Eng., F.I.E.E., F.I.E.R.E.*

SUMMARY

The object of this paper is to trace the development of the semiconductor industry itself, taking due account of economic and market factors, as well as the effects of technological change.

The historical growth of the industry is described, with particular emphasis on the structural changes which stemmed from the invention of the planar process and the development of integrated circuits. The technological, economic and market trends which are likely to dominate the growth of the industry in the 'seventies are discussed at some length, with considerable attention being given to the integrated circuit sector, in which the rate of expansion is certain to be most rapid and where the orderly acceleration of business activity will be disrupted to a degree by the strength of the innovative forces acting.

The many and diverse influences acting on the semiconductor industry are then analysed to enable forecasts to be made of changes in the future structure of the industry and of its expansion to 1980 in terms of global sales and employment.

* *Mackintosh Consultants Company Limited, Flemington Road, Queensway, Glenrothes, Fife KY7 5QL.*

1 Introduction

With the benefit of some historical perspective it is now possible to recognize the mid-point of the 20th century as something of a watershed in the evolution of industrial technology generally. The fact that the second half of the century is evidently different, in some quite fundamental way, has elicited various attempts to categorize it by means of labels such as the Space Age, or the Age of Automation, or the Computer Age. But the real point is that almost all of these new 'science-based' industries are basically dependent, often to a major degree, on the economic advantages and technical virtuosity of electronics technology, and to this extent a label more representative of the generic technology by which this period might best be characterized would be the Electronics Era.

This pervasive influence of electronics is due almost entirely to a single, cataclysmic event—the invention of the transistor. That event sparked off an astonishing quarter-century of research and development effort from which has stemmed a family of semiconductor devices of enormous variety, capable of carrying out almost all basic electronic functions (other than at the extremes of high power and/or frequency) more cheaply, more reliably and at higher levels of complexity than was dreamed possible at mid-century. Thus it is the semiconductor industry which has played the pivotal role in revolutionizing the cost, performance and reliability of an enormous range of electronic systems, and it is to the study of this industry, and of the trends which may influence its future structure, that this paper is devoted.

Forecasting anything—be it technological trends, the development of new products and markets, changes in economic or cost factors, or the structure of industries—is basically a matter of analysing historical and current trends (and the reasons for them) and then, by utilizing the most relevant analytical techniques, assessing the probable effect of all identifiable, relevant and significant forces of change, and making a reasoned judgement about the most probable future trend pattern. For this reason, the paper begins by briefly outlining how the semiconductor industry has developed so far, goes on to discuss the technological, market and economic trends which will dominate its development in the 'seventies, and finally offers my personal opinion of the structural changes which I believe will have taken place by 1980.

2 Historical Development of the Semiconductor Industry

As is well known, there was rapid worldwide appreciation of the importance of the work of Bardeen, Brattain and Shockley and it was only a matter of about five years from 1948 before simple germanium devices were appearing on the market in commercial quantities. During the remainder of the 1950s the industry grew at a very rapid rate, the acceleration being fuelled in the main by the development of a controllable and high-yield process for manufacturing germanium alloy transistors in large quantities. By the end of the decade, therefore, a substantial semiconductor industry already existed, based firmly on germanium technology, and it is interesting to reflect on the fact that, probably due to the liberal licencing policies of the Bell Telephone Laboratories, there was really no monopoly of semiconductor competence in the United States. In fact, in Japan a viable transistor industry was very rapidly established, and in Europe Philips was outstanding in terms of both size and technical excellence.

But meanwhile, back in California, a development was taking place which was to revolutionize this still fledgling industry. Many of the companies which quickly recognized the importance of this new invention went on to much fame and fortune; most of those that couldn't, or wouldn't, accept that a wholly new technology had arrived—to the inevitable and rapid decline in the importance of germanium—either withered away or in some cases are still struggling to recover the ground lost while clinging too long to an obsolescent technology. I refer, of course, to the development of silicon technology in general (with diffusion and epitaxial techniques being contributed primarily—again—by Bell), and to the invention of Fairchild's planar process in particular.

From 1960 onwards the dominating influence on the semiconductor industry was the remarkable speed with which these new silicon techniques were mastered, extended and exploited so that by today, only about a decade later, the 'semiconductor industry' mainly implies that industry which is based on silicon technology—with the bulk of it, except at the high power end, being planar.

The 1960s was also the period during which the US established a dominating position in the semiconductor field generally, and it is helpful in attempting to forecast how the industry will develop in the future to analyse a few of the salient reasons for this domination.

Clearly, the most important factor in providing the US technological lead was the tremendous R & D funding provided throughout the 1960s by the US government in its search for more reliable, higher performance and smaller electronic systems for space and defence purposes. The exact value of this expenditure is difficult to estimate (for obvious reasons) but probably amounted to between £150M and £200M cumulatively between 1961 and 1970, and it is important to note that the bulk of these funds was channelled to industry, not the universities or the government research laboratories. The result was that, as and when the possibility arose of a commercial 'fall-out' from a government-funded semiconductor R & D programme, industry was usually well placed technically to exploit the situation. It is also important to realize that in many cases the R & D support from the US government agencies was on a clear understanding that commercial fall-out was not only possible but desirable.

A related but subsidiary factor was the existence in the US of several industrial research and development laboratories of very high quality, which were able to provide a fertile environment for these government-funded programmes. The creation of these laboratories (mainly during the early 1950s) was very much an act of faith by US industry in the value of R & D, influenced no doubt by the very favourable impressions gained of the work of the Bell Telephone Laboratories, in particular, before and during the Second World War. Their creation was also strongly and actively supported by official circles in the US who could see that they would develop into an important national asset.

The third major advantage enjoyed during this critical period by the US semiconductor industry was (and still is) the existence of a very substantial domestic market for its products. Thus, in the early 1960s, not only was there a strong demand for military and space systems, but in the case of the major industrial customer—the computer industry—something in excess of 90% of the Western World's production was in the United States. This meant that in the early, critical days of the industrial applications of silicon devices, and integrated circuits, especially, the US industry benefited from having some customers who were receptive to innovation, wealthy enough to pay high prices for development samples and sophisticated enough in their knowledge of technology, economics and value engineering to be able to appreciate the long-term benefits of the new technology.

It should also be noted that the US was blessed with a breed of talented and technically-informed entrepreneurs, backed up by a sophisticated, powerful and knowledgeable financial community, so that once the implications of this new technology became evident, the way was clear for an enormous and relatively well-managed investment in productive capability to take place.

The achievement by the US of technological and commercial ascendency in this vital new industry was not wholly due, of course, to its own skill and good fortune, but was abetted to a large degree by the shortcomings of America's main industrial competitors. In general terms, the situation outside the US was one where managements and governments failed to recognize the importance of the new silicon technology, where the human resources to develop it were not available, and where—in most cases—the markets were not technically sophisticated enough to use its newer products. The net result was that by the end of the decade many of the US (silicon) companies were firmly established and had penetrated the free world semiconductor markets to an overwhelming degree, reinforced in many cases by their creation of indigenous production facilities in various strategic locations. Apart from the area of power devices, Western Europe's current demand for silicon devices, especially integrated circuits, is primarily met from production sources controlled by American capital, and only in the case of Japan has this US domination been resisted to any significant degree, primarily by the vigorous intervention of the Japanese Ministry of Industry & Trade (MITI). Even then, the Japanese semiconductor industry lags, technically and commercially, well behind the US, as evidenced by the preponderance of US-made m.o.s. i.c.s currently used in Japanese electronic calculators and similar equipments.

3 Dominant Trends in the 'Seventies

3.1 *Technological Trends*

If there is one single characteristic which distinguishes the semiconductor industry from most others it is the astonishing rate and persistence of technical innovation, which has gone on now for 25 years and which has carried the industry from the point-contact transistor to present-day, complex l.s.i. products, embracing on the way a variety of devices ranging from the photo-transistor to high-power s.c.r.s. All of the evidence points to the fact that, as in the past, the principal force of change in the semiconductor industry will continue to be technical innovation, and that the new products and processes which result will have a profound influence on the future structure of the industry.

The salient features of the present state of the technology can be summarized as follows:

(a) Silicon is probably the best understood material in common industrial use, and is readily and cheaply available at incredible levels of purity and crystalline perfection.

(b) The technology of most power and discrete devices is relatively static, but this is far from true in the case of i.c.s.

(c) The planar process—an elegant and sophisticated amalgam of precision photo-lithography and complex high-temperature chemistry—is the best example of a true mass-production (batch) technique anywhere in the electronics industry, and is constantly being improved and extended.

(d) In the area of assembly and packaging techniques, considerable improvements have been effected, both for discrete devices and for i.c.s.

(e) Today's bonding and chip packaging techniques are generally cheaper and more reliable, and significant increases in operator productivity have been achieved through the use of mechanized assembly methods.

So far as future developments are concerned, the most important point to make is that there is no possibility during this decade of a significant displacement of silicon by any other type of semiconductor. Clearly, other materials (e.g. compound semiconductors) will begin to carve out new markets for themselves during the decade but these will be relatively small and will be mainly in applications beyond silicon's ability to service. For example, this decade will certainly see a very substantial growth in the market for solid-state, light-emitting diodes and displays, but this will tend to reinforce, not erode, the market for silicon devices, e.g. control (access) circuits.

In the processing and assembly of semiconductor devices most of the technological innovation has been in that sector of the industry in which the growth has been most rapid, i.e. integrated circuits. Since most of the development effort will continue to be focused on i.c.s the remainder of this section is confined to that area.

The principal processing technology will remain a combination of diffusion and photolithography (i.e. basically planar), and the progress will tend to remain evolutionary rather than revolutionary. The bulk of i.c.s used in 1980 will clearly resemble those made today, utilizing the same material and most of the same basic processes. Chips will contain more active devices and will be more densely packed; systems will be partitioned to waste less silicon area on bonding pads, and chips will require less lead-bonding labour and, in many cases, will become amenable to complete mechanization of lead-bonding, assembly and packaging. Through a combination of circuit development and process improvements, circuits will become basically simpler, and performance factors such as speed and power consumption will be improved considerably.

In both the m.o.s. and bipolar areas, technological progress will come through general improvements in processing, such as the use of projection printing instead of in-contact masks, and the automation of human-intensive portions of the process such as mask-alignment. In m.o.s. specifically, quite dramatic improvements in performance will stem from the newer techniques such as silicon (instead of metal) gates, greater use of complementary m.o.s. circuits, and improvements in speed through the use of deposited silicon films on insulating substrates and, possibly, ion implantation techniques.

In my view, the dominant trend of process development in the m.o.s. area will be to bring together these distinctive new processes so that their individual benefits are combined to yield m.o.s. circuits of extremely high performance as measured by today's standards. The main impact of devices made by such methods will be in memories and in applications (e.g. watches and calculators) where high packing densities, cheapness and low power dissipation are salient requirements.

During the decade we shall also see the development of new devices based on, but significantly different from, the m.o.s. principle. One such is the m.n.o.s. 'reprogrammable' memory device; another, of much greater potential importance, is the charge-coupled device (c.c.d.) which is essentially an analogue shift register with applications for pulse compression in communications equipment and in the longer term as a very cheap, serial-read-out memory element.

In the bipolar area, there will be the continuing development of new processes such as the collector-diffused isolation (c.d.i.) and isoplanar techniques. It is unlikely, however, that devices based on these new processes will make a large impact in the open market since, despite promising an improvement in costs over conventional bipolar devices, their simplified structure leads to a degradation in performance to the point where these devices will be competing against m.o.s., the fundamental advantages of which are now becoming quite overwhelming.

Although likely to be challenged strongly by new forms of m.o.s. structure, it is very likely, on the basis of current evidence, that bipolar devices will continue to dominate the high-speed end of the market. The level of single-chip integration, however, will be determined by the yield improvements which can be achieved in this relatively complex bipolar processing, rather than by the development of discretionary wiring, which I believe to be a basically uneconomic process due to the poor utilization of silicon area and high testing costs.

In general terms, the trend of silicon technology will be to develop along more or less well established lines, with refinements of technique leading to ever-improved process control and yields to the point where very large chips, possibly containing as many as 10^5 to 10^6 individual circuit elements, will be produced routinely, cheaply and in very large numbers. When this point is reached, as it almost certainly will be during the present decade, highly complex electronic functions will be available at a very modest cost and it will be entirely appropriate for systems designers to begin treating such 'massive' i.c.s as a basic ('raw') material.

The effects of having available such 'massed' electronics at very low cost will indeed be revolutionary, and it is probably no exaggeration to claim that there is little understanding yet of the full impact which micro-electronics technology will eventually have on industry and society, and that this is a revolution which has hardly yet begun.

Finally, in this brief summary of technological trends, it is appropriate to say a word about the application of computer-aided-design techniques to i.c.s. The present situation, without doubt, is that relatively little real use is being made of c.a.d. methods, although considerable amounts of money are being spent on developing viable systems. On the other hand, the use of computer-aided techniques is bound to increase significantly as the complexity of l.s.i. designs extends beyond the point at which the (unaided) designer can no longer cope efficiently. The economics of i.c. design and the future split of the market between custom and catalogue devices will depend critically on the outcome of this work.

3.2 *Market Trends*

In general terms, the markets for small-signal discrete devices will continue to diminish as the changeover to i.c.s gathers further momentum, although this will not apply, of course, to special devices (microwave, opto-electronic etc.) which do not suffer from competition with i.c.s and for which the markets will generally show impressive growth. For power and/or very high-frequency devices the markets will continue to grow at a substantial if unexceptional rate as these devices extend their range of performance and find their way into a number of new applications. The main growth of the semiconductor market, however, will be in i.c.s so that it is again appropriate to focus most attention in this area.

If the 1960s could be described as the decade in which digital i.c. logic 'came of age', the 1970s will be the decade in which the use of i.c. memory functions will enjoy an explosive growth. The two basic reasons for this are that software problems are becoming increasingly intractable and expensive to solve, and that all hardware devices, logic and memory, are becoming very cheap. Thus, increasing emphasis in the design of computing systems will go into expanding the hardware capability in order to simplifiy the software. In addition, present-day computers tend to be 'under-memoried' in relation to their purely arithmetic capabilities due to the fact that it has so far been much cheaper to provide a lot of logic than a lot of memory.

Integrated circuit memory devices will capture the bulk of this greatly-expanded new market for computer random-access memories, now served predominantly by ferrite cores. Although there are currently strong indications of significant design activity using bipolar circuits it is my own view that the bulk of the future main-frame memory market will be m.o.s. Many American and some European m.o.s. companies are now heavily committed to m.o.s. memories and the volume of innovative effort, plus the competitive market situation, will ensure, I believe, that m.o.s. will become the cost-effective solution for most applications.

There are also indications that substantial inroads will be made by semiconductors into the market for rotating memory devices such as disks and drums. In one disk/drum memory area the crossover between the prices of semiconductor and rotating machinery is already approaching. At the present time, fixed-head systems with full read-and-write electronics on every head, such as are used in display systems, run between 0·4 and 0·6 cent/bit, whereas one US company is reported willing to quote 0·3 to 0·4 cent/bit in quantities of several hundred million bits per annum. Although the penetration by i.c. memories of this market is obviously more speculative and further off than in the case of main-frame memories, the example quoted does emphasize the very competitive position now being achieved by i.c. memories, and underlines the prediction that the future market for these semiconductor devices is going to be very large indeed, despite the inevitable resistance which will be put up by the magnetic device industry.

The market for i.c.s represented by the defence and aerospace industries has always been of great importance and, of course, has been responsible for spawning many new products which have then contributed to the growth of other markets. It is quite clear that, on a worldwide basis, this (basically military) market will continue to grow at a substantial rate and have an important but indirect effect in other areas.

Telecommunications has long held promise of being a natural and dominant market for i.c.s but has so far failed to develop to a significant degree largely because of the difficulties which are encountered in making electronic exchanges and related equipment compatible with the very large, existing investment in electro-mechanical exchanges. These problems are gradually being overcome and, together with the developments of p.c.m. systems and data communications etc., mean that the worldwide telecommunications market for i.c.s will grow very rapidly during the 1970s, although probably not as fast in the UK as in many other countries.

The foregoing market trends are already discernible, and are quantitatively predictable with a reasonable degree of certainty. There are, however, several other i.c. market categories which will undoubtedly show rapid growth, albeit from very small beginnings, during the decade, but which are very difficult to predict with any high degree of accuracy. The main markets in this category are electronic calculators, consumer electronics (including clocks and watches, and possibly some form of video recording) and automotive electronics.

One of the impressive features of the semiconductor industry has been the way in which, through its own product innovation, it has created its own new markets. This is well exemplified by electronic calculators, the design of which has been revolutionized over the past three years or so by the development of appropriate, complex m.o.s. i.c.s. This market now represents a very significant outlet for American m.o.s. companies, and it is clear that this strong trend away from electromechanical calculators will continue.

Somewhat more speculative is the possible development of battery-operated and portable calculators which are cheap enough to ensure very widespread use in any field of activity–industrial, commercial or domestic—in which relatively simple (e.g. 'slide rule' and 'adding machine') calculations need to be carried out. In my view, there is no basic reason why such a device cannot eventually be made at a cost equivalent to a relatively simple transistor radio and, at about £10 each, possibly containing an i.c. priced at between 50 and 100p, could represent a very substantial worldwide market.

Consumer electronics as a whole also has very great potential as a major i.c. market but is difficult to predict at this point in time with any accuracy. For example, an m.o.s.-controlled, numeric read-out electronic watch is already on the market in the United States, albeit at a high price. If, however, the price of such watches can be reduced to the level (£20, say) where a mass market opens up, the worldwide market for 'timing' i.c.s would also become very significant. I believe that such a development will indeed take place which, together with the many other, well-known potential applications of i.c.s in consumer electronics (especially in television sets), will make this one of the most important market categories by the late 1970s. This belief is reinforced by my prediction that it will be on consumer electronics that much of the innovative talent in the worldwide semiconductor industry will be turned.

In the field of automotive electronics, the potential market for semiconductor devices in general, and i.c.s in particular, is very large indeed and this trend will undoubtedly be reinforced by the current trend in many countries towards greater legislation over antipollution and safety standards. In brief, there is already substantial use of semiconductor devices in voltage regulators, tachometers etc. and, in due course, this is likely to be extended to a wide variety of vehicle subsystems in the general fields of safety and convenience (e.g. electronic locks and sobriety testers, anti-skid braking controls and climate, window and seat controls), engine controls (e.g. electronic ignition, fuel injection, anti-collision controls), performance indication (speedometer, instrumentation and sensors, self-diagnostic systems etc.), power distribution (e.g. generators) and entertainment (radio and tapes).

There is also the question of whether there will be a strong trend towards the preferential use of custom-designed i.c.s instead of standard, catalogue devices. There are many manufacturers who are gambling on such a trend developing, and it will be difficult to justify some of the current expenditure on c.a.d. unless they are right, although, as stated earlier, the use of computer-aided techniques is now becoming essential in the design of many of the more complex i.c.s. On the other hand, the whole history of the semiconductor industry has been that yesterday's custom product is available today 'off the shelf'.

In my view, there will be a definite trend towards a greater number of i.c. designs. In part this belief is based on the fact that the more complex the product the greater the room for tailoring it to specific customer requirements. In the second place, the heavy current investment in c.a.d. may be a self-fulfilling wish in the sense that the design facilities must be made to 'pay', therefore the design services will be sold cheaply enough to encourage widespread use, and therefore many customers will indeed go to customized designs. Finally, of course, the development of the technology is likely to be such that there will be very little difference in the processing cost between 'custom' and 'standard' slices. However, the special testing requirements and the necessary amortization of the design costs, will inevitably make custom i.c.s somewhat more expensive unless made at production volumes similar to standard products, and it is therefore probable that many of these new designs will soon find their way into the catalogues of standard products.

3.3 *Economic Trends*

In most sectors of the semiconductor industry the business climate has been almost continuously influenced by the phenomenon of long and rapid price reductions—averaging out for semiconductor devices as a whole at about 30% per annum. The causes, quite obviously, have stemmed mainly from improved process control (i.e. higher yields) and from severe business competition.

The effects on the costs of a whole range of electronic equipment have been astonishing. It is sometimes difficult to appreciate the full impact of this semiconductor price revolution on the user industries and it helps, I think, to reflect on the fact that only six or seven years ago logic gates used in discrete-component computers were being assembled at production costs of between £2·00 and £5·00, whereas today's i.c. equivalent (considerably improved in performance, moreover) can be purchased for around 10p, representing an average price reduction of about 40% per annum. In the same vein, during the early-to-mid 1960s some forms of silicon transistors were exhibiting price reductions as high as 50% per annum as the markets, production volume and competition all grew at a rapid rate.

The effect on the semiconductor industry of these remarkable price reductions has not by any means been wholly bad. Although they have made life difficult for all producers, and especially so for those unable to match the production volumes (and, therefore, low costs) of the market leaders, it has been the very rapid price reductions which have been the main cause of the explosion of semiconductor usage. In this sense, the i.c. industry in particular can be said to have made its own opportunities and, by imaginative pricing policies, has captured a very substantial part of the total available electronic components market, as well as creating completely new markets, in a surprisingly short space of time.

So far as the future is concerned it seems clear that the dominant economic trend will be a continuing drive towards lower costs. In general, this drive will be manifest in two completely different ways. So far as discrete small-signal and power devices and 'classical' i.c.s (i.e. small-scale integration, or s.s.i., devices) are concerned, it will take the form of maximizing production volume and productivity and of minimizing package and labour costs. Some efforts will be made to decrease chip sizes and to improve 'front-end' (i.e. slice processing) yields but it is now a fact that the savings possible in this area have become quite limited. This is not the case for complex (e.g., l.s.i.) i.c.s, however, for which the cost-reduction drive will mainly be concentrated on die size, yields and 'front-end' costs generally, since in this case assembly labour is only a small part of the total cost.

Unless political forces intervene, the trend towards increasing use of cheap, 'off-shore' labour will certainly continue for semiconductor devices involving significantly labour-intensive operations (e.g. wire-bonding), although this decade will probably see the extension to s.s.i. devices of the fully mechanized assembly and packaging techniques currently used in the production of many small-signal discrete devices.

Within the semiconductor industry itself there seems little prospect of any real easement of the fierce competition which has already been alluded to. The average rate of profitability for most of the US industry, and for substantially all of the industry outside the US, has been highly unsatisfactory for a very long time. This basic weakness of the semiconductor industry became a chronic condition early in the 1960s, and during that decade its full implications have been camouflaged for many US companies by two important factors. The first, mainly applicable to the established companies, was the very sizeable flow of US government funds into the industry, for the main purpose of supporting a substantial part of the R & D programme.

The second factor, of immense value to new and old US companies alike, has been the exciting investment image which all science-based industries, and not least semiconductors, have enjoyed in the US. The result has been a remarkable and sustained willingness by American investors—institutional and private—to provide the capital for investment in the provision of production facilities etc., despite the evident inability of many of these companies to finance their own expansion by means of profits earned. In particular, the high-risk venture capital organizations have provided the major source of start-up funds for most 'spin-offs', and this process would not have worked to anything like the same degree in the absence of the US over-the-counter market for dealing in 'unquoted' shares.

Outside the US, and in Europe particularly, there has been very little protection from these chilly blasts of competition from US companies, many of whom have enjoyed the benefits of economies of scale to a high degree. Most European semiconductor manufacturers, other than those confined to the power device sector, have struggled along, subsidizing their losses from profits made in other product sectors and/or limiting the magnitude of these losses by keeping their semiconductor activities relatively small. Unfortunately, this kind of policy is practically guaranteed to maintain the state of non-profitability.

In my view this prolonged profits squeeze puts the industry into a basically unstable situation. It is clearly dangerous to the long-term health of the industry when two major international semiconductor companies, one on each side of the Atlantic, can both sustain substantial losses on their semiconductor activities in 1970 and 1971. If the basic cause of this situation is the severity of the competition, then one important force for change in the future will be the search to find ways to reduce this. Whether, from the broadest point of view, this would be either possible or desirable is, I think, a moot point.

4 The Future Structure of the Industry

The picture which the international semiconductor industry presents today can be summarized briefly, as follows. The part of the industry which is concerned mainly with producing power devices is relatively stable and profitable and enjoying steady if unspectacular growth. The small-signal discrete and 'classical' i.c. sectors, however, are just emerging from a difficult period of ferocious competition, vicious price wars and chronically poor financial returns. At the innovative end of the i.c. industry, however, things look brighter, with new technologies and new products being developed —especially in the m.o.s. area—and new companies being spawned (mainly in the US) to exploit them. There is little shortage of American venture finance for such companies at the present time.

Starting from a consideration of the present situation, and allowing for the forces of change and dominant trends which have been discussed above, it is possible to predict with some confidence the future structure of the industry, especially so far as the US is concerned. Europe is more difficult, mainly because the present insecure foundations of the British and Continental companies will serve them poorly as launching pads into the 1970s. Detailed and accurate predictions about other parts of the world cannot be dealt with in this paper.

Dealing first with the non-i.c. sectors of the industry, I believe that the power and small-signal discrete areas will continue along basically well-mapped paths. The power device sector will progressively become more international in character, but with many of the major participants—in Europe, especially—still relying heavily on in-house 'captive' markets. Growth rates will be coupled quite closely to, although in many cases markedly higher than, the growth of G.N.P. in each of the nations in which each individual company trades to a significant degree. Companies deeply committed to small-signal discretes will suffer a period of extremely difficult trading conditions, with markets contracting and competition expanding. Towards the end of the decade, however, it is probable that the number of participants will have drastically fallen, that prices will harden and that it will then become possible for the survivors to trade profitably.

Turning to the i.c. sector, it is clear that in the US new companies will continue to be formed, especially in m.o.s., but probably at a lower rate than in recent years. A few of the well-managed and well-financed new producers will have a good chance of becoming major participants in the industry, particularly those in the m.o.s. and memory areas, but some of the recently-formed companies will merge together or be taken over.

With one exception, it is unlikely that any of the new groups will operate at the classical end of the i.c. product range, although new participants in both linear and non-saturated bipolar logic i.c.s are likely as these markets grow. The exception will arise if and when one of the larger companies manages to penetrate the high-volume end of the business through the successful development of a fully mechanized assembly and packaging technique, and if this does happen it will have the important subsidiary effect of reversing the trend towards greater and greater use of off-shore labour.

In general, the overall structure of the US industry, and the general nature of the forces acting on it, are not likely to change drastically over the decade. There will still be cycles of famine and glut, under-capacity and over-production. New price wars will flare up and lead to the collapse of the most tenuous participants, and the gradual evolution of new technologies and new products will give continuing but fewer opportunities for newcomers to challenge the established companies.

Predicting the future structure of the European i.c. industry is a very much more difficult matter, however. The only aspect which seems incontrovertible is that the US subsidiaries will continue to claim a substantial fraction of the total available European market. Whether that fraction will be greater or less than it is today depends mainly on the ability of the European i.c. producers to survive and/or succeed and, to a secondary degree, on any actions taken by individual governments to protect and succour their native i.c. producers.

Taking all relevant factors into account, and based on a fair degree of personal awareness of the individual circumstances of many semiconductor companies, American and European, I believe that of approximately eight substantial European i.c. producers operating today, not more than three or four will survive in a significant form until 1980. Some operations will be reduced to the status of in-house facilities (of dubious value) or will merge in an attempt to cut losses through economies of scale, and some will simply cease operations.

Clearly much needs to be done, by both European governments and industrialists, to foster the development of i.c. companies which possess the right combination of determination to succeed, financial 'muscle-power' and willingness to learn what it takes to succeed in this difficult industry. Only when effective actions are taken by such companies to work out and put into effect new strategies relevant to the 1970s will it be possible to predict with any confidence the eventual emergence of viable and competitive European participants in the integrated circuits industry.

5 The Scope for Growth

Finally, it is, I think, appropriate to conclude with some brief comments on the impact of semiconductor devices on the growth of equipment industries, and with some personal estimates of the global size and importance of the semiconductor industry, so that the implications of the invention of the transistor can be appreciated in economic, as well as scientific and technological terms.

In the first place it is important that the critical role played by semiconductor devices in stimulating the growth of electronic capital goods and consumer industries should be fully appreciated. The rapid development, for example, of the entire computer industry has been almost entirely due to the gradual development of ever cheaper, more reliable and lower power semiconductor logic circuits, and the defence and aerospace industries have been revolutionized by the availability of light-weight and high-reliability semiconductor components covering a very wide range of electronic functions. In addition, many consumer electronic products (particularly devices such as transistor radios, portable televisions and, most recently, cheap electronic calculators), the manufacture of which represents a substantial constituent of the economies of several nations in the Far East, would simply not exist today were it not for the success of the semiconductor industry in evolving techniques to produce enormous numbers of devices at costs which, judged by the standards of 25 years ago, seem incredibly low.

Turning to the economic importance of the semiconductor industry itself, by my calculations the value of the annual production output of the industry on a worldwide basis currently stands at about £1300M (assuming the Communist bloc to be producing about 25% of the output of the OECD countries), and this probably represents a total, worldwide *direct* employment of between 250,000 and 350,000. (In addition, of course, a very large amount of indirect employment is generated by the semiconductor industry, just like any other major sector of the economy.) By 1980 I estimate that the value of total world production output will have increased by a factor of almost three, to a global figure of about £3800M, and that total direct semiconductor employment will by then have grown to approximately 600,000.

In brief it is my basic contention that, enormous though its impact has already been, the semiconductor device has a still greater role to play in the further growth—in scope and in depth—of electronics technology generally and that the ultimate boundaries, economic and technical, of semiconductor technology lie well beyond the bounds of established predictive methods. When it is appreciated that all of this stems from a few relatively crude experiments, allied to some superb theoretical and deductive skills a mere 25 years ago, it is clear to me, at least, that the Nobel Prizes awarded to Messrs. Bardeen, Brattain and Shockley, for their joint invention of the transistor, were richly deserved indeed.

6 Acknowledgments

Some parts of this paper stem from analyses conducted as part of a major study of the microelectronics industry recently carried out by the author's company, and thanks are due to the sponsors of that project, the National Research Development Corporation and the Department of Trade & Industry, for permission to use some of that material. The author also wishes to express his gratitude to Group Captain A. Brown, Dr. M. J. Turner and Mr. G. A. White for helpful discussions during the preparation of this paper.

Manuscript received by the Institution on 6th November 1972. (Paper No. 1507/MT 2.)

Contributors to this issue

Professor J. J. Bates is Head of the Electrical Engineering Branch at the Royal Military College of Science, Shrivenham. He received his B.Sc. degree in 1948 from Birmingham University, and after apprenticeships with the English Electric Co. and the British Electricity Authority (now the Central Electricity Generating Board), he returned to Birmingham in 1950 to do postgraduate work on temperature rises in electric machines; he received his Ph.D. degree for this research. After a short period with the British Iron and Steel Research Association, he joined the Royal Military College of Science, Shrivenham, in 1954 as a lecturer in the Electrical Engineering Branch. In 1959 he moved to Imperial College London, as a lecturer in the Electrical Engineering Department, specializing in electric machines and electric traction. In 1965 he became the first Reader in Electrical Engineering Science at the University of Durham. He was appointed to his present position at Shrivenham in 1967. His principal research interests are in the physical limitations of d.c. machines, particularly those relating to commutation.

Mr. J. G. Davies is a Senior Technical Officer and head of the electronics section of the Physics Department at St. George's Hospital, London, where he has been employed since 1955. He served with the R.A.F. as a pilot and as a flight engineer from 1943 to 1948 and, prior to going to St. George's, he was at the Institute of Cardiology. His involvement in pacemakers started in 1955 and he received an M.B.E. in recognition of his pioneer work in this field in 1966.

Dr. K. J. Dean (Fellow 1965, Member 1952) has been Principal of South East London Technical College since September 1972. Previously he was for four years Vice-Principal at Twickenham College of Technology and his earlier teaching posts were at Letchworth, Kingston and Norwood. Dr. Dean is a graduate of London University and his higher degrees were based on theses in physics, mathematics and electronic engineering. His main research interests have been logic design and special purpose computing and he has written numerous books, papers and articles on circuit theory and techniques and integrated circuits and digital instruments.

Dr. Dean has been active in Institution affairs and is currently a member of the Papers Committee and Chairman of the Computer Group Committee; he has also chaired or served on the Components and Circuits Group Committee and on Conference Organizing Committees.

Dr. J. C. Brice read natural science at the University of Cambridge and, after graduating in 1956, joined Mullard Research Laboratories where he is currently Head of the Crystal Growth Department. He was awarded his doctorate in 1969 for his research into the preparation of materials of interest to the electronics industry. He is the author of two books and numerous papers.

Mr. Carl S. den Brinker was born and educated in the Netherlands and after service in the Royal Dutch Navy as a signals officer he settled in the UK in 1952. In 1956 he joined British Telecommunications Research in Taplow, ultimately becoming head of the circuit development section. In 1962 he moved to Texas Instruments, where his present position is that of Manager, New Business Development. Mr. den Brinker is the author of several papers and holds 20 patents. He is a member of the Research Board of Letchworth College of Technology.

Mr. J. S. Brothers (Member 1971) joined the Plessey Company's Allen Clark Research Centre in November 1961. He initially worked on aspects of silicon integrated circuit technology, moving to integrated circuit design in 1965. Since this time he has been involved in analogue and digital i.c. design, and in the application of computer-aided design to integrated circuits; he is author of several papers in these areas. At present he is a Senior Principal Scientist and holds the position of Integrated Circuits Design Manager.

Dr. R. E. Colyer received his undergraduate and postgraduate training at the University of Bristol where he obtained a B.Sc. degree in 1958 and a Ph.D. in 1963 for research on speed-changing induction motors. From 1963 to 1969 he was an Assistant Professor at the Middle East Technical University in Ankara, Turkey and on returning to the UK spent a year as a Research Associate at the University of Bristol. Since 1970 he has been a Senior Lecturer at the Royal Military College of Science where his main interests are the application of semiconductor devices to the control of electrical power and machinery.

Mr. P. E. K. Donaldson served in the Royal Navy from 1944 to 1947 and on demobilization went to Cambridge University, where he read electricity and aeronautics, obtaining his degree in 1950. From 1954 to 1958 he worked at the Physiological Laboratory, Cambridge as a Technical Officer and subsequently as a Senior Technical Officer. During this period he started on visual prosthesis research with Professor G. S. Brindley, with whom he moved to the Maudsley Hospital, South London, to form the MRC Neurological Prosthesis Unit, and he is now a member of the Scientific Staff of the Medical Research Council. Mr. Donaldson is author of 19 papers on engineering subjects, and one book; he was Editor of *Medical and Biological Research* from 1963 to 1968. He was recently elected to the Council of the I.E.E.

Mr. C. C. Fielding graduated in physics at the University of Durham in 1948. He spent a year carrying out research with the British Scientific Instruments Research Association at Chislehurst, Kent, and then joined the Radar Research and Development Establishment at Malvern, Worcestershire. He continued with the Royal Radar Establishment at the same location until 1964 when he was appointed Assistant Director of Electronics Development (Ground) in the Ministry of Aviation and subsequently Ministry of Technology. In 1968 he returned to RRE as Head of the Electronics Group.

Professor William Gosling (Fellow 1968), who graduated from Imperial College, University of London, in 1953, has been at the University College of Swansea since 1958. He was appointed to the Chair of Electrical Engineering in 1966 and has recently relinquished the headship of the Department of Electronic and Electrical Engineering to undertake a three-year term of office as Vice-Principal of the College.

Professor Gosling has written over 20 papers and four books, his principal research interests being in instrumentation and field effect devices. He has been closely concerned with IERE affairs for several years and is currently Chairman of the Professional Activities Committee. Following two years as a member of the Council, he has now been elected a Vice-President.

Professor H. L. Hartnagel (Fellow 1972) who graduated Dipl.Ing. in electronics from the Technical University of Aachen, Germany, in 1959, worked for Telefunken for a short period and then joined the Institut National des Sciences Appliquées, Lyons, France. In 1961 he took up an appointment as a research fellow in the Department of Electronic and Electrical Engineering at the University of Sheffield, subsequently becoming a lecturer, in 1968 a senior lecturer, and a reader in 1970. He received his Ph.D. from the University of Sheffield in 1964 for a thesis on electron beam behaviour and the D.Eng. degree in 1971. Two years ago Professor Hartnagel was appointed to the new chair in electronic engineering at the University of Newcastle upon Tyne. He is author of a book 'Semiconductor Plasma Instabilities' and of numerous papers for one of which, on Gunn diode ultra-fast logic, he received the Institution's Charles Babbage Award.

Dr. J. B. Izatt studied electrical engineering at Aberdeen University, obtaining his B.Sc. in 1957. He remained at Aberdeen to do research on frequency modulation and received his doctorate in 1959. From 1960 to 1965 he was with the BBC Research Department and worked on aerials, transmission lines and modulation systems; the latter included work on compatible single side-band amplitude modulation. He then joined the Solartron Electronic Group as leader of the Product Evaluation Group, followed by three years in charge of oscilloscope design. His present position, to which he was appointed in January 1972, is that of Group Leader of Dynamic Analysis Research and Development.

Mr. Bruce A. Joyce graduated in chemistry at the University of Birmingham in 1956 and then served for two years as a Technical Signals Officer in the Royal Air Force. He joined the Allen Clark Research Centre of the Plessey Company in 1958, before joining Mullard Research Laboratories up his present appointment in 1969. His fields of interest relate to semiconductor thin films and surface properties, and he has published over thirty papers on these topics. He is a founder committee member of the British Association of Crystal Growth, and a committee member of the Thin Films and Surfaces Group of the Institute of Physics.

Dr. Ian M. Mackintosh (Fellow 1964) is Chairman and Managing Director of Mackintosh Consultants Company Ltd., an international firm of consultants which he formed in 1968 to specialize in providing services to the electronics and allied industries. His involvement with the semiconductor industry goes back about 20 years to his original research work on III-V compounds. This was followed in 1956 by six years with Bell Telephone Laboratories, at Murray Hill, New Jersey; it was during that period that Dr. Mackintosh developed the first four-layer (p-n-p-n) triode, precursor to the silicon controlled rectifier (s.c.r.) and published the first theory of this device. Later, he was at Pittsburgh for about two years as a Department Manager in the Westinghouse Central Research Laboratories. After returning from the United States in 1964, Dr. Mackintosh became General Manager of Elliott-Automation Microelectronics Ltd.

Dr. Mackintosh is a graduate of the University of Nottingham where he obtained his degree in physics in 1953 and a Ph.D. in 1956. He has published over 25 major scientific papers and he delivered the I.Prod.E.'s 1967 Nuffield Memorial Lecture on Microelectronics. He is on numerous professional and academic committees, and served as the industrial adviser to the UK delegation to the OECD Study of the Technology Gap. Dr. Mackintosh also serves in a part-time capacity as a Visiting Industrial Professor of the University of Edinburgh, and is a Council Member of the Council for National Academic Awards.

Mr. A. Martinez (Fellow 1969, Member 1959, Graduate 1953) obtained his baccalaureate at Murcia University, Spain and then studied at E.M.I. College from 1950 to 1953, to gain the City and Guilds full technological certificate in telecommunications engineering. He worked at Peto Scott on the development of radio and television receivers for three years and in 1956 joined the Marconi Company, where he was concerned with the development of ruggedized industrial television cameras, three-vidicon colour cameras and camera control units. He became a group section leader in the broadcasting division working on colour systems and was responsible for a transistorization programme in encoders.

In 1963 Mr. Martinez joined the Perdio Company and was appointed chief engineer of the Sunderland plant, where the first all-transistor television portable receiver was produced in this country. After two years as an independent consultant he joined the British Radio Corporation and in 1968 was appointed chief engineer of the Bradford plant, which is one of the largest in Europe.

Dr. J. B. G. Roberts graduated in physics at the University College of Swansea in 1958 and gained a Ph.D. there in 1962 for work on electrical and thermal conduction in solids. Since then he has been at the Royal Radar Establishment, Malvern, where he has worked mainly on plasma physics, ionospheric radar scattering and signal processing.

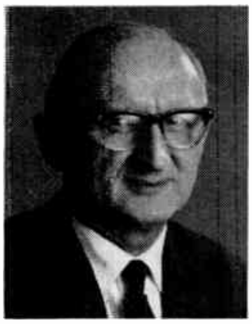
Mr. G. G. Scarrott is now Manager of the Research and Advanced Development Organization of International Computers Limited in Stevenage. After the war he worked for some seven years at the Cavendish Laboratory, Cambridge, on nuclear physics instrumentation, and in 1953 he joined the Computer Department of Ferranti Limited, which now forms part of ICL. Mr. Scarrott has published papers on electronics for nuclear physics, including a pulse height analyser, and, since moving into the computer field, on wire-type acoustic delay lines, slave stores, efficient use of logical devices and principles for system design.

Dr. A. A. Shepherd is a graduate in physics of the University of Manchester, where he also took his Ph.D. degree, carrying out research on oxide-coated cathodes, and in 1950 he joined the lecturing staff of the Physics Department of the University of Keele. In 1954 he moved to the Wythenshawe Laboratories of Ferranti to initiate research on semiconductor devices in the Electronic Components Division, subsequently becoming Chief Engineer of the division. In 1967 he was appointed General Manager of the Instrument Department of Ferranti, a post he held until September 1970, when he was appointed to his present position of General Manager of the Electronic Components Division. He is the author of numerous papers and articles, including three previous papers in the IERE Journal.

Dr. M. Smollett graduated in physics at the Royal College of Science, London, where he also gained his Doctorate with a thesis on lattice dynamics in 1951. He then studied the behaviour of polarized electrons at the Sorbonne, Paris.

He joined the Mullard Company at their Research Laboratories in 1953, working initially on infra-red detectors and semiconductors. He left the Research Laboratories in 1956 to help set up the first Mullard semiconductor plant in Southampton and he is now Chief Development Engineer for semiconductors.

Professor John J. Sparkes graduated in physics from Manchester University in 1948. After a short period in the Admiralty Signals Establishment the joined British Telecommunications Research in 1952 and became Head of the Physics Section. For ten years he was closely associated with all aspects of transistor development. In 1962 he went to the Electrical Engineering Department at Imperial College as a Senior Lecturer in Communications and in 1967 he was appointed a Reader at the University of Essex. In the following year he was awarded a Ph.D. as a result of his publications on transistors. In 1970 he became Professor of Electronic Design and Communications at the Open University. Professor Sparkes's recent research interests are in the fields of artificial intelligence and pattern recognition.



Dr. J. R. Tillman graduated in physics with 1st Class Honours and as a Prizeman from Imperial College, University of London, in 1932; he subsequently carried out research in electron difraction and was awarded a Ph.D. He then moved to nuclear physics research and worked with P. B. Moon on the discovery of the thermal neutron. Dr. Tillman entered the Post Office Research Department in 1936 and was seconded to the Telecommunications Research Establishment, Malvern from 1943 to 1944. He was promoted to Senior Principal Scientific Officer in 1952 with responsibilities particularly for active electronic devices, and he was closely concerned with the development of special transistors for submarine telephone cable repeaters. He was awarded a D.Sc. by London University in 1955. Dr. Tillman was appointed Deputy Director of Research of the Post Office in 1965 and his responsibilities include materials, devices, microelectronics, postal automation, etc. He has served on many IEE and Institute of Physics committees, and is a Visiting Professor to the Electrical and Electronic Engineering Department of City University.

IERE News and Commentary

The 47th Annual General Meeting of the IERE

The Institution's 47th Annual General Meeting (the 11th since Incorporation by Royal Charter) was held at the London School of Hygiene and Tropical Medicine on Thursday, 7th December 1972.

The President, Mr. A. A. Dyson, opened the meeting at 6 p.m. when sixty-eight corporate members had signed the Attendance Register. Before proceeding to the formal business of the meeting, Mr. Dyson said that the Director and Secretary, Mr. G. D. Clifford, was unfortunately unable to be present because of illness. This was the first Annual General Meeting at which Mr. Clifford had not been present in thirty-four years, and all members warmly joined the President in wishing him a speedy recovery.

The President then called on Mr. R. C. Slater (Deputy Secretary) to confirm that all members had received due notice of the meeting. Mr. Slater reported that, in accordance with the Bye-laws of the Institution, Notice of the Eleventh Annual Meeting of the Institution since its Incorporation by Royal Charter, together with the Agenda, were published on page S.136 of the September 1972 issue of the Institution's Journal, *The Radio and Electronic Engineer*.

Minutes of previous Annual General Meeting

The Deputy Secretary reported that the 46th Annual General Meeting, the 10th since incorporation by Royal Charter, was held on 2nd December 1971 and the minutes of that meeting were published in the January 1972 issue of *The Radio and Electronic Engineer* (page S.1).

The President stated that no comment had been received on the minutes and he therefore moved that this record of the Annual General Meeting be approved, to which members gave unanimous agreement.

Annual Report of Council

After referring members to the Annual Report published on pages S.161–181 of the November 1972 issue of *The Radio and Electronic Engineer*, Mr. Dyson said:

'The Industrial Relations Act has laid down the principle that all employees, as well as shareholders, have a right to know about their Company's affairs. While this does not apply to the Institution in the same way as to a manufacturing or commercial organization, it is of great advantage for all of us to know what the Institution is doing, its problems and perhaps its failings! Certainly the Report before us tonight covers all the Institution's activities and fairly comments on our problems.

'I believe the unity of the engineering profession to be one of our most important goals if we are to achieve our rightful status in the technologically-based world of today and the future. Two aspects of this Institution's endeavour in this direction are detailed in the Report of the Executive Committee which has devoted a considerable amount of time in evolving, with the other Chartered Engineering Institutions, a common policy towards registration under the Industrial Relations Act and the promulgation of a meaningful code of Professional Conduct for Chartered Engineers. The other important step towards unity is the registration of Technician Engineers. Chartered Engineers and Technician Engineers are complementary: one cannot work effectively without the other and it is therefore, right that the greatest co-operation and unity should be engendered between them. As a result of the setting up, through CEI, of the Technician Engineer Section of the Engineers Registration Board, it has been possible to invite nearly 700 members of this Institution, mainly in the grade of Associate, to have their names included on the Register so that they might use the title 'Technician Engineer (CEI)' and the designatory letters 'T.Eng. (CEI)'. I regard this not only as an important move towards the unity of the engineering profession in this country but also as an essential step towards the unification of engineers and engineering qualifications throughout the European Economic Community.

Photograph above. The President signs the Minutes of the previous Annual General Meeting, watched by the Deputy Secretary, Mr. R. C. Slater.

'The Executive Committee has also been deeply involved in the revision of the Institution's Bye-Laws, and these were the subject of a Special General Meeting held last September.*

'Although the Institution no longer conducts its own Examinations, the Education and Training Committee and the Examinations Committee have a continuing role to play. Indeed, in a rapidly advancing technology the importance of their role is likely to increase rather than decline, and it is incumbent upon them to ensure that the stipulated academic and training requirements are adequate not only for the present, but that they will provide an engineer with a sound basis for advancement throughout his professional life. I think you will agree that the Annual Report shows that they are executing their remit with considerable skill both for the future of the Institution and the community which we all serve.

'We are like any other organization in that we cannot stand still—we must either grow or decline! I am happy to say that once again the Membership Committee Report shows a healthy state of growth. The figure for overall membership has not increased markedly, but this has been caused by the removal from the Register of a large number of Students who clearly would not be able to attain the required academic standards. Against this, however, there has been a significant increase in the number of Corporate Members.

'It is stated in our Royal Charter that the object of the Institution is to promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering. Thus it gives me particular pleasure to read the reports of the Professional Activities Committee and the Papers Committee. The former records that over 160 lecture meetings, colloquia and conferences were held throughout the country during the year, and this in itself is a tribute to the many members who freely give of their time to sit on Specialized Groups, Local Section and Conference Committees. The Report of the Papers Committee gives some indication of the enormous amount of work which goes into collecting and assessing the papers which makes the Institution's Journal, *The Radio and Electronic Engineer*, so widely and authoritatively quoted throughout the world.

'Our Treasurer will shortly be reporting on the Institution's Accounts and I will not encroach on his domain, except to pay tribute to the work of the Finance Committee in combating the escalating costs while at the same time making provision for increased services to members.

'Before formally moving the adoption of the Annual Report I would like to pay tribute to the many members who have supported me by their services on Council and Committees during the year, and I would particularly like to thank our Director and Secretary, Graham Clifford, for the enormous help he has given in piloting me through sometimes difficult and turbulent waters'.

The President then invited comment on the report, and none being forthcoming, formally moved its adoption, which was unanimously approved.

Auditors' Report, Accounts and Balance Sheet

Before asking Mr. George A. Taylor to present the accounts, Mr. Dyson said that the Institution had always been fortunate in its choice of Honorary Treasurers, but none had served so long as Mr. Taylor. However, being now retired, Mr. Taylor had said that he thought that this should be the last year in which he should accept election as Treasurer. The President was sure that all members would like to express their appreciation for Mr. Taylor's long and valuable service, especially in these last few years when inflation had brought financial problems to this and every Institution.

Mr. Taylor thanked the President and members for their expression of confidence, and said that he was conscious of the honour of being re-elected Honorary Treasurer of the Institution for another year, during which he would continue to serve the Institution. He felt, however, that it was time the office passed to a younger man, and that when he retired at the next Annual General Meeting it might be appropriate for him to say a few words on the financial development of the Institution over the years.

The Accounts and Balance Sheet were presented by Mr. G. A. Taylor, Honorary Treasurer.

Turning to his immediate task in presenting the Income and Expenditure Accounts for the year ended 31st March 1972 and the Balance Sheet at that date, Mr. Taylor said:

'I am not going to weary you with fine detail but I will content myself by touching upon one or two salient points which are perhaps relevant to your consideration.

'Once again we see the continuing steep rise in costs reflected in the Expenditure Account, for example administration has risen by £8000, or just about 10%, and the increase in this item alone exceeds the total increase in revenue from subscriptions!

'The cost of publishing and distributing the Journal rose by £4500, but from this we can deduct over £2000 from increased sales of the Journal.

'Subscriptions to CEI have risen by about £250 because of increased membership and we shall of course have to pay more as our membership increases each year. On this item it is worth noting that our association with CEI now costs us just over £3000 a year, to which must be added a net drop in revenue of £1500 by loss of examination fees which are now the concern of CEI. It would seem therefore, that the price we have to pay in trying to achieve academic unity with other constituent bodies is around £4500 per annum. We must continue to hope that this will ultimately prove to be a profitable investment.

'It will be seeen that the surplus of income over expenditure is now down to just over £1000 compared with £9000 for last year. A surplus of this slender magnitude cannot of course be regarded as auguring well for the future. It is obvious that if the current rate of inflation is maintained, this kind of surplus will disappear rapidly into an adverse balance.

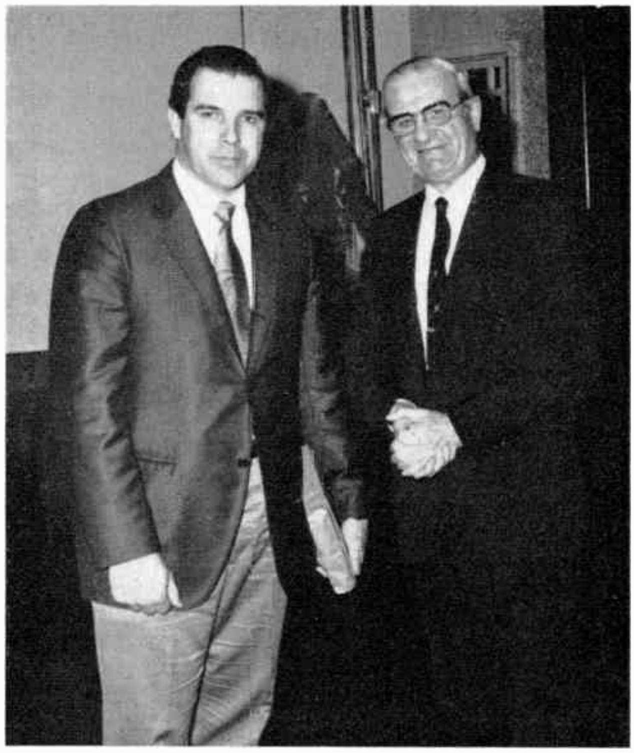
'If now we refer to the Balance Sheet, we will see that our accumulated deficit has decreased by about £2000, but it is

* Reported in *The Radio and Electronic Engineer* for October 1972. (pp. S.149–50).

Before the start of the A.G.M. members discuss points in the Report. Two Vice Presidents, Mr. J. Bilbrough and Air Commodore S. M. Davidson, can be seen on the left of the picture, while another member of Council, Mr. R. Larry, is in the centre of the lower row of seats.

The President with Professor D. S. Campbell, who received the Clerk Maxwell Premium, and the Institution's Editor, Mr. F. W. Sharp.

Professor D. R. Towill and Dr. P. A. Payne receive their J. Langham Thompson Premium.



Mr. Dyson and Mr. R. G. Drever pause for the photographer's record.

The President has just presented Mr. M. H. Lee (*left*) and Dr. V. J. Phillips with their Premium.

going to take a long time before we are able to extinguish the whole of the deficit and to begin to build any reserve.

'We have added about £6500 to our fixed assets in respect of necessary alterations and repairs to 8 and 9 Bedford Square.

'The market value of our investments has increased over the year by over £5000 and in these days this must be regarded as highly satisfactory, particularly when, during the year, we have realized about £800 on the sale and re-purchase of stock and in addition received some £600 in dividends and interest.

'What are the conclusions to be drawn from the accounts before us? I am afraid they are painfully obvious—as are the remedies. If we are to remain solvent, our costs must be held down even more severely and economies must be made. Revenue will need to be re-appraised, an accelerated rate of increase of new members is necessary and, as a last resort only, a further review of subscription rates must be made, although I know Council are most reluctant to consider revised rates unless and until they are convinced there is no other alternative.

'Although I am informed that so far as we know at present Value Added Tax is not likely to have any devastating effect on our operating costs, to nurture the thought that a change in our tax system will result in any reduction would be, as Dr. Johnson once said, the triumph of hope over experience! We have never yet seen this happen.

'Mr. President, Ladies and Gentlemen, I know that Council will give consideration to the few points I have tried to make. I can assure you that your Finance Committee is already engaged in a careful review of the situation and are preparing proposals for the attention of Council. I beg to move that the Income and Expenditure Account for the year ended 31st March 1972 and the Balance Sheet as at that date be adopted by this meeting.'

Mr. Dyson thanked Mr. Taylor for his lucid presentation of the accounts and invited a member to second their adoption. Of the numerous members wishing to second the proposal, Mr. S. R. Wilkins (Fellow) was recorded, and members then unanimously approved the adoption of the Auditors' Report, Accounts and Balance Sheet.

Election of Council for 1972–73

The President called upon the Deputy Secretary to refer to the list of nominations for election to the Council, and Mr. Slater reported that there were no opposing nominations to those made by Council and circulated to corporate members by publication in the September 1972 issue of *The Radio and Electronic Engineer*, and a ballot had therefore not been necessary. (A full list of the 1973 Council is published on page (i) of this issue.)

Mr. Dyson thanked the membership for confirming his own re-election as President, saying that he was conscious of the honour and assuring members of his interest in the welfare of the Institution and of his continued promise to do all he could in its interests. Mr. Dyson then referred to the fact that all members of Council must retire in rotation after serving the statutory periods in the various offices, and continued: 'I would particularly like to thank Mr. F. N. G. Leevers, Dr. I. Maddock and Mr. A. S. Pudner who have completed their term as Vice-Presidents. In addition to their services as Vice-Presidents all three have served the Institution for many years, either as members of Chairmen of the various Standing Committees.

'In their places we have elected Air Commodore S. M. Davidson, Professor W. Gosling, Mr. S. R. Wilkins and Mr. A. St. Johnston; and have re-elected, so that they also serve a maximum period, Mr. J. Bilbrough, Professor W. Gambling and Group Captain C. K Street.

'We also particularly welcome as Ordinary Members of Council Mr. P. A. Allaway and Brigadier R. Knowles (Fellows) and Dr. J. R. James and Professor D. E. N. Davies (Members); at the same time thanking Mr. R. C. Hills and Captain P. J. Poll for their services on Council.

'Finally we greatly welcome the election of Professor H. E. M. Barlow, a distinguished Engineer and an Honorary Fellow of the Institution.'

Appointment of Auditors and Solicitors

The President said that he would like to combine Items 5 and 6 of the Agenda, namely the appointment of Auditors and of Solicitors. Both having served the Institution well for over thirty years, Mr. Dyson asked for approval to the re-appointment of Gladstone, Jenkins and Company as the Institution's Auditors, their remuneration to be at the discretion of Council, and of Braund and Hill as Solicitors to the Institution.

The President's motion was carried unanimously.

Presentation for Premiums and Awards

After the conclusion of the formal business of the Meeting the President passed to the presentation of Premiums and Awards to the authors of outstanding papers published in the Institution's Journal during the year. Ten papers were considered to be of sufficient merit for recognition in this way and in only one instance were the authors unable to be present to receive the books or scientific instruments which they had chosen.

Mr. Dyson asked the Institution's Editor, Mr. F. W. Sharp, to announce the names of the prize winners and their papers, details of which were given in an Appendix to the Annual Report. The recipient of the Clerk Maxwell Premium for the most outstanding paper on a subject published during 1971 was Professor D. S. Campbell who now occupies the chair in electrical engineering at Loughborough University of Technology but was in industry when he contributed his paper on 'Electrolytic capacitors'.

Two of the Premium winners, Professor D. R. Towill and Dr. P. A. Payne received the J. Langham Thompson Premium for the third and second occasion respectively for their joint paper on automatic testing of control systems.

Although the Institution often singles out papers by overseas authors for recognition in this way, it is seldom that an author is able to attend the meeting in London to receive his prize. It was therefore particularly pleasant for the President to be able to welcome Mr. Robert G. Drever of the Woods Hole Oceanographic Institution and to present him with the A. F. Bulgin Premium for Measurements for the paper which he had jointly contributed with Dr. Thomas Sanford at the 1970 Conference on Electronic Engineering in Oceanography. The paper dealt with a free-fall ocean current meter using geomagnetic induction.

The meeting closed at 6.40 p.m. and was immediately followed by an Address on 'The Electronic and Radio Engineer in the EEC' by Mr. J. P. Engels, Chairman of Philips Electronic and Associated Industries.

The MacRobert Award consists of £25,000 and a Gold Medal, and is given annually for an outstanding contribution by way of innovation in engineering or the physical technologies or in the application of the physical sciences, which has enhanced, or will enhance, UK prestige and prosperity.

The Award was instituted by the MacRobert Trusts in 1968 with the aim of honouring individuals or small teams of individuals. The MacRobert Trusts were founded by Lady MacRobert of Douneside and Cromar, widow of Sir Alexander MacRobert, Bt., founder of the British India Corporation.

The obverse of the MacRobert Gold Medal. The reverse bears the MacRobert Armorial Bearings; a photograph was published in the March 1972 issue of *The Radio and Electronic Engineer*. Brief details of the Award winning invention are given on the rim. The Medal, which is 75 mm in diameter, was designed by Mr. Arnold Machin, O.B.E., R.A.

1972 MacRobert Award
for the Invention of a New X-Ray Technique

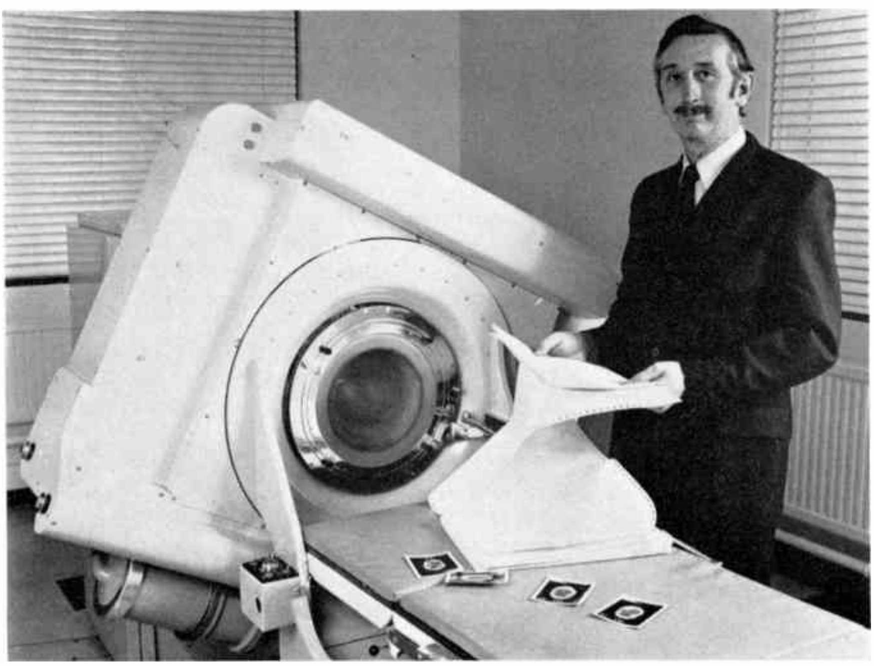
Mr. Godfrey Hounsfield and EMI Limited have won the 1972 MacRobert Award for the invention of a new X-ray technique. This technique is the basis of a new computerized system for diagnosing brain disease. Mr. Hounsfield is Head of the Medical Systems Section of EMI's Central Research Laboratories at Hayes, Middlesex.

The MacRobert Award, often described as the Nobel Prize for engineering, consists of a Gold Medal and £25,000 prize money. It is presented annually, in recognition of the technological innovation contributing most significantly to the prestige and prosperity of the United Kingdom. HRH Prince Philip presented the Award cheque to Mr. Hounsfield at a ceremony at the Savoy Hotel, London, on 22nd November and the Gold Medal, presented to EMI Limited, was received by Sir Joseph Lockwood, Chairman of EMI.

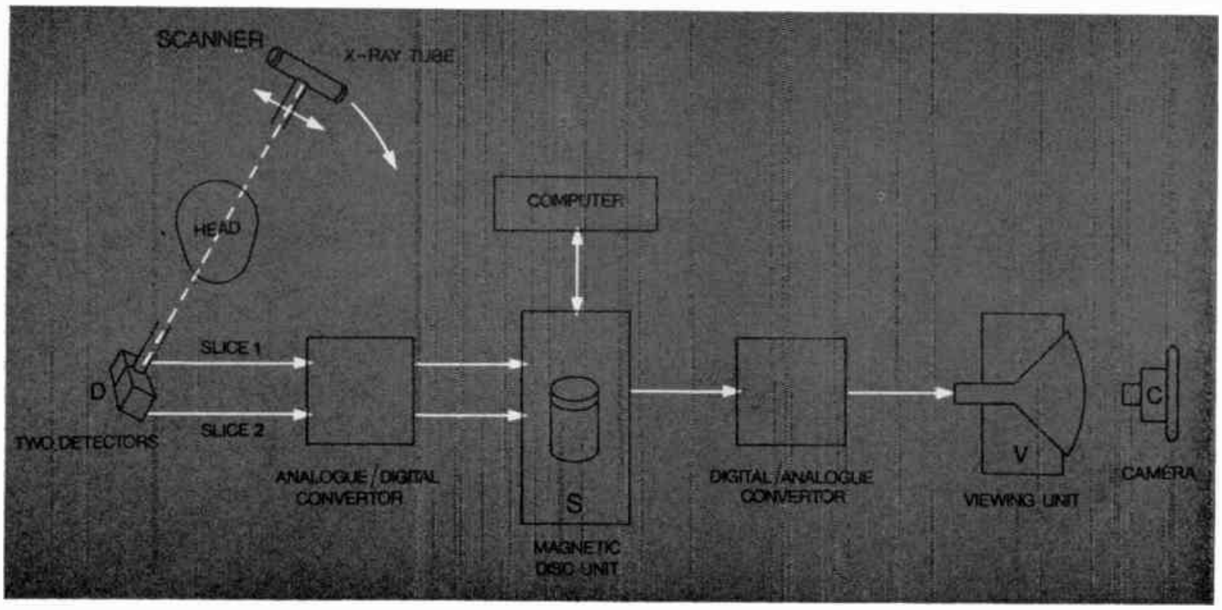
Assessment of the Invention

In his citation for the Award, the Chairman of the MacRobert Award Evaluation Committee, Lord Hinton of Bankside, said:

'One of the medical referees consulted during the evaluation stated that no comparable discovery has been made in this



Mr. Godfrey Hounsfield, inventor of the new X-ray technique which has won for him and EMI Limited the 1972 MacRobert Award. He is standing beside the patient-scanning unit of the EMI-SCANNER. The patient lies on the table with his head inside the opening which is surrounded by a rubber bag containing water. This keeps the head stationary and in the correct position while the scanner rotates.



Schematic of the EMI-SCANNER system

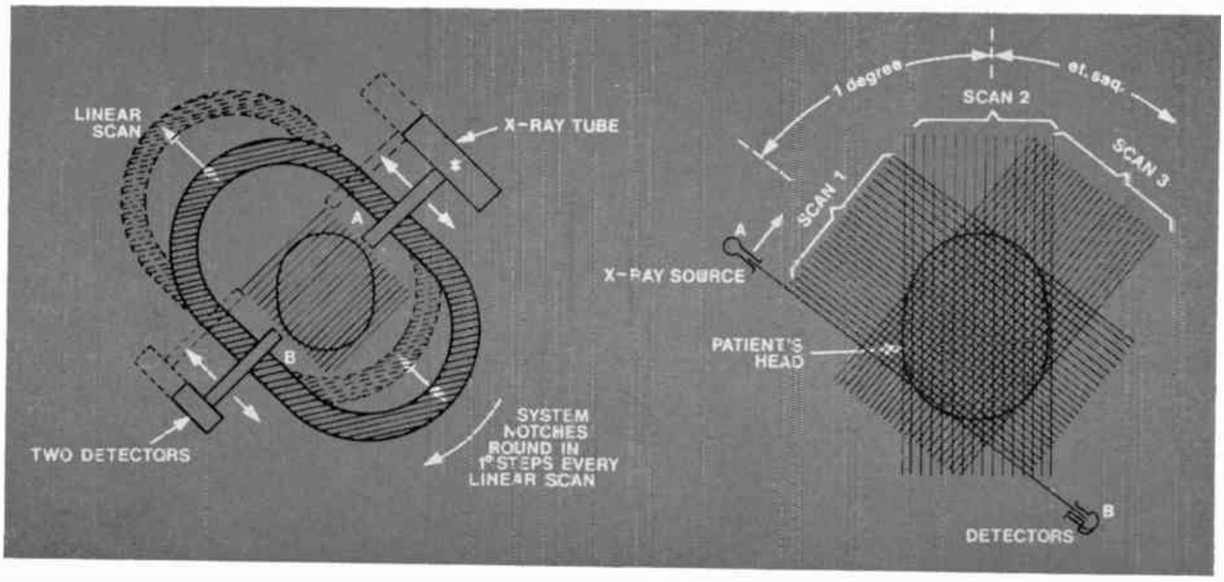
field since Röntgen discovered X-rays in 1895, and we agree with him. The EMI-SCANNER system developed by Mr. Hounsfield within the Central Research Laboratories of EMI is epoch-making, because it breaks away from the photographic techniques for recording X-ray pictures which in principle have remained unchanged since Röntgen's day. These techniques have the inherent defect that they seek to show a three-dimensional object in a two-dimensional picture, without the benefit of perspective. This confuses the information recorded on the film. Also, details of healthy and diseased brain tissue, which have low absorptive characteristics, are masked in X-ray pictures by the surrounding, denser bone of the skull.

'As well as overcoming these obstacles to the diagnosis of disease in the brain, the EMI system avoids the need, associated with other diagnostic techniques, for injecting the patient with substances such as radio-opaque fluids, which are not without risk to the patient. With this new system, the patient is not put at risk, and is no more inconvenienced than by a chest X-ray. Without the need for unpleasant injections, 100 times more information is extracted from the X-ray photons than is possible using conventional means.

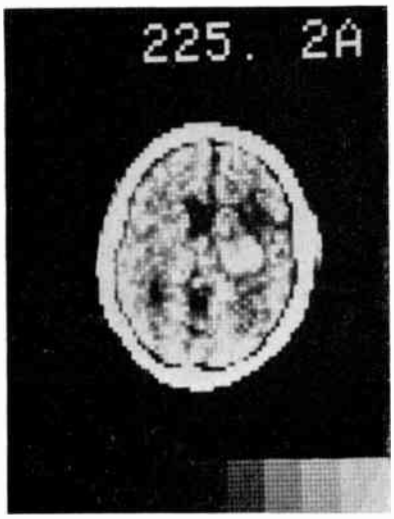
'The technique has, as yet, been applied only to cranial examination but the MacRobert Award Evaluation Committee believes that the principle of this invention can be widely used—not only in medical diagnosis but perhaps also in the examination of inanimate objects.

'There is another aspect of EMI-SCANNER which is remarkable. In these modern days it is rarely that one finds great developments which are the work of one man. The EMI-SCANNER is different: the submission for the MacRobert Award was prepared not by the inventor, but by EMI, who stated, "Mr. Hounsfield has been the guiding expert throughout all aspects of the work. The EMI-SCANNER was as much a one-man invention as anything can be these days".'

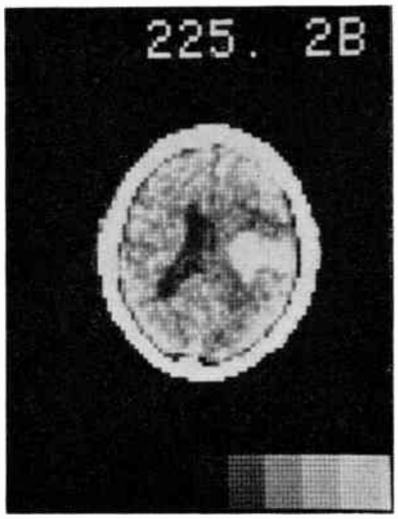
Since the EMI-SCANNER was introduced earlier this year, orders have been received totalling nearly £600,000, of which



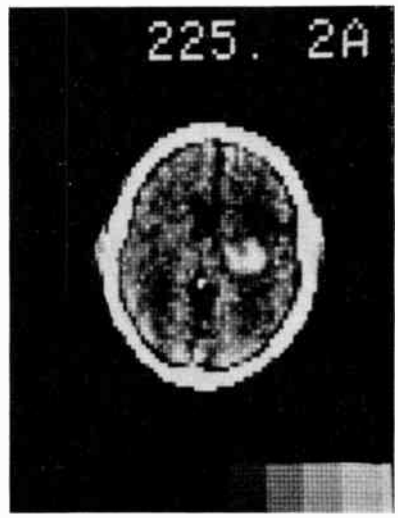
The scanning method



Haematoma. Tomographic slice taken between 5·7 cm and 7·0 cm above the orbito-meatal line.



Same patient. Slice taken between 7·0 cm and 8·3 cm above the orbito-meatal line.



As first picture but with a different setting on the 'window width' and 'window level' controls to isolate features of specific interest.

Typical pictures produced by the EMI SCANNER.

£400,000 worth have come from the USA. The first installation is at Atkinson Morley's Hospital in Wimbledon, South West London, whose medical staff worked closely with Mr. Hounsfield in developing the equipment.

How it works

The EMI-SCANNER is a system of computerized axial tomography. The patient's brain is examined as a series of layers (either 8 mm or 13 mm in thickness) by a scanning unit housing very sensitive X-ray detectors. These record the X-ray photons passing through the brain. In four minutes, during which time the fully-conscious patient relaxes on the examination table, the scanning unit is rotated around the patient's head in 180 one-degree steps.

At each step, 160 accurate readings are taken of brain tissue absorption characteristics from a narrow beam of X-rays. The readings are digitized and stored in a magnetic disk store. The information is then fed to a small computer which solves 28,800 simultaneous equations representing the readings taken from each layer. From these calculations it produces a picture on a cathode-ray tube, made up as a matrix of 6400 picture points. Each 3 mm square point represents the X-ray absorption coefficient of the brain tissue at that point in the patient's head, calculated to 0·5% accuracy.

The picture becomes available approximately 6 minutes after the patient has been scanned. It may be studied on the c.r.t. screen immediately, or may be recorded photographically for later examination. The results may also be printed out as a pattern of numbers, to give accurate and detailed information on the brain tissue at each point in the patient's head. This provides an important alternative aid to the diagnosis of certain conditions.

As well as providing 100 times more information on brain tissue than is possible by conventional means, the EMI system avoids the need for skilled medical staff to be in attendance and eliminates factors which makes it necessary for patients to spend time in hospital either before or after a brain examination. Thus it not only represents a major advance in diagnosis, but it provides important benefits to the patient and to hospital administrations.

On 12th April next, under the auspices of CEI, Mr. Hounsfield will give the MacRobert Award Lecture describing his invention. The Lecture will take place at the Institution of Electrical Engineers, Savoy Place, London, W.C.2; further details will be announced later.

Submissions for the 1973 MacRobert Award are now invited and entries should reach CEI by 30th April 1973. The rules and conditions of the Award, which were summarized in the March 1972 issue of The Radio and Electronic Engineer, can be obtained from: The MacRobert Award Office, Council of Engineering Institutions, 2 Little Smith Street, London SW1 3DL (Telephone 01-799 3912).

Members' Appointments

The Council has congratulated the following members of the Institution whose names appear in Her Majesty's New Year's Honours List published on 1st January 1973:

APPOINTMENTS TO THE MOST EXCELLENT ORDER OF THE BRITISH EMPIRE

COMMANDER, CIVIL DIVISION (C.B.E.)

Mr. Percy Albert Allaway (Fellow 1971). (Mr. Allaway is Chairman of EMI Electronics Ltd; he joined the Company as Design Engineer in 1939 and was appointed to the Board in 1954, becoming Managing Director in 1961. A fuller note of Mr. Allaway's career was published in the September 1972 issue of *The Radio and Electronic Engineer* in connexion with his nomination for election as a member of the Institution's Council.)

OFFICER, CIVIL DIVISION (O.B.E.)

Mr. John William Cheesbrough (Member 1956). (Mr. Cheesbrough is Regional Engineer with the Midlands Telecommunication Region of the Post Office. He has been with the Post Office for over 48 years and for much of this time was concerned with the design of automatic telephone exchanges within the Midlands Region.)

MEMBER, CIVIL DIVISION (M.B.E.)

Mr. Philip James Darby (Member 1969). (Mr. Darby is Head of Technical Quality Control with the Independent Broadcasting Authority, a position he has held since 1967. He joined the BBC as a Grade C Engineer in 1950 and moved to the Authority when it was formed in 1955. Before appointment to his present post he was for eight years Engineer-in-Charge of the ITA Station at Dover.)

CORPORATE MEMBERS

Mr. E. J. Chappell (Member 1965) has recently returned to the Ministry of Defence as a Professional and Technology Officer Grade 1 at the Admiralty Underwater Weapons Establishment, having spent three years on Project Control at the European Space Research Organization's Technology Centre at Noordwijk, Holland.

Mr. A. B. Close (Member 1971, Graduate 1967) who was recently awarded the Master of Science degree in Systems Engineering by the University of Surrey, has been appointed a Senior Scientific Officer at the Ministry of Defence (Procurement Executive), Boscombe Down. Mr. Close was previously a Senior Engineer with Recording Designs Ltd.

Mr. Olawole Fatimilehin, M.Sc. (Member 1972, Graduate 1968), an Executive Engineer 1 with Nigerian External Telecommunications Ltd., was recently promoted to Senior Engineer, and is responsible for all services at Ikoyi High Frequency Receiving Station, Ikorudu Transmitting Station and the International Maintenance Centre, Lagos, Nigeria.

Mr. J. G. Coates (Member 1963), who was previously Marine Service Manager with Redifon Telecommunications Ltd., has been appointed Lecturer in Marine Electronics at Brunel Technical College, Bristol.

Lt. Cdr. B. Craig, RNZN (Member 1967, Graduate 1963) has recently left active service in the Royal New Zealand Navy and has taken up a post with the New Zealand Ministry of Transport, Civil Aviation Division, where he is a Senior Telecommunications Engineer in the Radar Section.

Mr. J. S. Dahele (Member 1971, Graduate 1968) has been granted two years' leave of absence from Standard Telecommunications Laboratories, Ltd., Harlow, to take up the appointment of Visiting Lecturer in the Department of Electronics, The Chinese University of Hong Kong.

Squadron Leader T. P. Dickins, RAF (Member 1970) has recently been appointed Elect.Eng.(Air), HQ 11 Group, RAF Bentley Priory, Stanmore. Sqd. Ldr. Dickins was formerly Officer Commanding Electrical Engineering Squadron, RAF Leuchars.

Mr. B. E. Dowden, B.Sc. (Member 1968) has been appointed Principal Lecturer in the Department of Technology, School of Engineering, Ipswich. He was previously on the staff of Wimbledon College of Technology.

Mr. J. C. Edwards (Member 1971, Graduate 1966) has recently formed his own company, Sunthorne Electronics Ltd., at Farnborough, Hampshire, and holds the post of Managing Director. He was formerly Director, Standfast Burglar Alarm Co. Ltd. and Technical Director, Pack's Infotel Ltd.

Mr. M. R. Green (Member 1963, Graduate 1961), a Project Manager with Marconi Space and Defence Systems Ltd., Dunfermline, has now been promoted Quality Manager with the Company.

Mr. D. W. Grierson (Member 1970, Graduate 1967), formerly Senior Electronic Engineer, Data/Telemetry, with Aerotrain Systems Inc., California, has now taken up the appointment of Electronic Engineer, Consultant, with H. L. Yoh, Consultants, a division of Day and Zimmerman Inc., Hollywood, California.

Mr. D. H. J. Mason (Member 1969) has been appointed Professional and Technology Officer, with the Department of the Environment in London. He was for three years a Senior Engineer, with Plessey Company at Poole, prior to which he was for some eight years with Hawker Siddeley Dynamics, Industrial Automation Division, where he was Test Engineer, Project Engineer NCB contracts, and latterly Project Engineer, Medical.

Mr. R. J. Montgomery (Member 1969) has now rejoined Ferranti Ltd. as Senior Integrated Circuit Design Engineer, at Wythenshawe, Manchester. Mr. Montgomery, formerly a Senior Electronics Engineer with Ferranti Ltd. at Bracknell, left them in 1971 to take up the post of Senior Engineer, Passive Components, with Waycom Ltd., Bracknell.

Mr. J. F. Pengelly (Member 1972, Student 1969) has been appointed Development Engineer, Underwater and Weapons Department, Marconi Space and Defence Systems Ltd., Portsmouth. He was previously a Senior Research Engineer with Plessey Telecommunications Research, Poole, having joined that Company in 1969 as a Research and Development Engineer.

Mr. A. M. Surti, M.Sc. (Member 1971) is now a Scientist on the staff of the Department of Scientific & Industrial Research, Applied Mathematics Division, Wellington, New Zealand. Before emigrating to New Zealand in 1971 he was a Senior Design Engineer with Plessey Co. Ltd., at Ilford Essex.

Mr. P. R. W. Webb (Member 1959, Graduate 1955) who returned to Canada last year to take up appointment in the Directorate of Technical Resources Management, Ottawa, as a Communications/Electronics Engineer in the Canadian Armed Forces, has now retired from the Service and is Standards and Metrication Officer with the Department of Industry, Trade and Commerce.

Mr. L. W. B. Worrall (Member 1971, Graduate 1968) who on his return from the USA in 1971 joined the Foxboro Yoxall Company, Redhill, as a Section Leader Methods Engineer, has been promoted Project Administrator—Nuclear Contracts.

NON-CORPORATE MEMBERS

Mr. A. K. Bagchi (Graduate 1970) has been appointed Lecturer in Mathematics, Mander College, Bedford. He was previously Telephone Traffic Superintendent with Post Office Telecommunications, London.

Captain R. N. A. Joy, B.Sc.(Eng.), REME (Graduate 1971), formerly with 6 Field Workshop REME, has been appointed Officer Commanding Parachute Squadron RAC Workshop REME, Candahar Barracks, Tidworth.

Mr. J. K. Murray (Associate 1959) has recently formed his own Company, JKM Consultants, at St. Albans, Herts. Until 1968 he held positions as a Technical Author and Design Engineer with Marconi Instruments Ltd. For the last four years he has been with Greenpar Engineering Ltd., first as Development Manager and, for a short period, Quality Control Manager.

CORRECTION

Mr. G. P. Bassom (Member 1969, Graduate 1964) has recently been promoted from Lecturer to Senior Lecturer in Applied Physics at Southampton College of Technology, and not to a Junior Lectureship as was incorrectly stated in the December 1972 *Journal*.

OBITUARY

The Council has learned with regret of the deaths of the following members:-

Advice has only recently been received of the death on 14th February 1972 of **Miss Yvonne Margaret Dyson Cooper, L.ès.Sc., B.Sc.** (Honorary Fellow 1941, Member 1927).

Miss Cooper was educated in France and was an accomplished linguist. Following the award of her Licencié ès Sciences in France, she returned to England and obtained a B.Sc. degree. She lived for the rest of her life in the Bournemouth area and in the 1920s held a senior appointment in the Test Department at an Army research establishment. Miss Cooper then took a teaching appointment and she was the Institution's Examiner for the Heat, Light and Sound and Electricity and Magnetism papers of the Graduateship Examination in the 'thirties. She served as an ordinary member of Council for a number of years until just before the War and her work for the Institution was recognized by election to Honorary Fellowship in 1941.

Although other commitments took up a large part of her later life, Miss Cooper continued to teach on a part-time basis at Bournemouth Technical College until only two years before her death at the age of 76.

Stephen Lascelles Morgan, M.C. (Member 1950, Student 1947) died on 19th September 1972, aged 53 years. He had been seriously ill since February.

A Londoner by birth, Mr. Morgan went to South Africa as a child and attended schools in Pretoria and Johannesburg. After three years as an Articled Clerk he served throughout the war with the South African Artillery and the Corps of Signals, mainly in North Africa and Italy, and was awarded the Military Cross.

In 1945 Mr. Morgan joined the South African Broadcasting Corporation as a Transmitter Engineer; he was appointed Supervisory Engineer in 1948 and Engineer-in-Charge of the Johannesburg Studios in 1950.

He joined Tramincor Industrial Electronics Ltd. of Johannesburg in 1961 and was a Director of that Company at the time of his death.

Lt. Col. Stanley Paton Morrison, Royal Signals (Member 1968) died on 13th August 1972, aged 41 years. He leaves a widow and two young daughters.

Educated at Morrison's Academy, Crieff, Lt. Col. Morrison attended the RMA, Sandhurst, and from 1951 to 1952 undertook Specialist Signal Officer training at the School of Signals. His early service included BAOR (1957–59), North Malaysia (1956–57) and Singapore (1957–59). He was then appointed to Eastern Command in the UK from 1959–62, followed by a year as Instructor at the Australian School of Signals. After successfully completing the Technical Staff Course at the RMCS Shrivenham, in 1965 he subsequently served as Commander of a Radio Squadron in 22 Signal Regiment. An excellent skier, he organized a winter warfare scheme in Norway for his Corps in 1966. From 1967 to 1969 he was seconded to the Ministry of Technology and for the next two years commanded the Malaysian Armed Forces Signal Regiment, being based in Kuala Lumpur. On his promotion to Lieutenant Colonel in June 1971 he was appointed to SHAPE, but after only three months he was taken ill and died a little under a year later in the Queen Alexandra Military Hospital.

Electronic Approach System for Berthing Tankers

A radar Doppler system which will provide an accurate indication of the closing speed of all tankers approaching the jetty has been ordered from Marconi Marine by the Mobil Oil Company for their Coryton Refinery terminal in the River Thames. Unlike earlier systems of its type, the Coryton installation will also give an indication of true distance of a vessel off the berth in addition to closing speed.

The system, which has become known as SAMI (Speed of Approach Measuring Instrument), was developed by Marconi Marine in conjunction with the Royal Radar Establishment at Malvern and the Esso Petroleum Company, and is intended primarily to protect oil jetties from excessive damage during the berthing of large tankers. The velocity of first impact on the jetty is critical and if the impact speed is more than about 7·5 m/min.—or about a quarter of a knot—there will be severe damage to both jetty and ship.

The pilot responsible for berthing the vessel is normally on the bridge wing some 30 m (100 ft) above the water and anything up to a quarter of a mile from the bows. The problem of estimating very low speeds from this position is, therefore, extremely difficult and the berthing operation is invariably a long and arduous task. SAMI was developed to help alleviate this problem by providing the pilot with information on the vessel's approach speed, and continuous range measurement during the final berthing operation.

The system makes use of the Doppler shift phenomena in a radar signal beamed at an approaching vessel from the jetty. The installation at the Coryton terminal will consist basically of two transmitter/receiver units operating in the 14GHz band, together with their associated 53 cm (21 in) dish aerials producing a 3° transmission beam. In addition there will be a control panel providing a meter reading of speed and range, information which is recorded simultaneously on a chart recorder to indicate overall changes in velocity and distance. A Gunn diode source generates the 14GHz signal. Balanced mixers compare the frequency of the transmitter and received signal, the difference, the Doppler shift, being converted to give a speed indication. For range, the Gunn diode is frequency modulated by a varactor diode, the frequency difference between transmitted and received signals being converted to a distance measurement.

The complete system is to be installed in the jetty and there will be no requirements for any on-board electronics other than a conventional v.h.f. radio link over which the information is passed to the pilot. The transmitter/receiver units will be spaced some 120 to 180 m (400 to 600 ft) apart at each end of the jetty with the control panel somewhere between.

The system is used during the two phases of berthing. The first phase is the channel approach during which the vessel is navigating the channel approaching the berth, the second, the final berthing manoeuvre during which the vessel is pushed sideways into the berth by tugs.

During the first phase the system is used to provide the pilot with details of his speed in knots down the channel, information he needs to enable him to stop the ship before it reaches the berth. To achieve this, the operator or observer stands behind the transmitter receiver nearest to the vessel and, using a sighting tube on the side of the equipment, steers the beam to follow the ship.

When the vessel is off the jetty, the second more critical phase of berthing begins. The system is now used to provide the pilot with highly accurate speed and range information of both the bow and stern sections of the vessel. The transmitter/receiver unit which has been used during the first phase and the second unit at the other end of the jetty are now both fixed at 90° to the line of the berth pointing straight out towards the vessel. Three speed ranges measure 0-50 ft/min, 0-100 ft/min, and 0-10 knots, while distance off can be measured up to 500 feet. During this phase, movement of as little as four feet per minute can be measured and distance off down to two feet.

Letters to the Editor

The Institution's Council does not necessarily agree with views expressed by correspondents.

Correspondence of a technical nature, or on any matter of interest to electronic and radio engineers, is welcomed.

From: Mr. R. B. Newton, B.Sc.
Mr. H. Pursey, B.Sc.
Mr. A. Gwyn Jones
and **Mr. L. H. Bedford,** C.B.E., M.A., C.Eng., F.I.E.R.E.

and **Mr. N. J. Keen,** B.Sc., C.Eng., F.I.E.R.E.

Electronic Aids for the Tuning of Musical Instruments

I was intrigued by Mr. L. H. Bedford's excellent article* which has increased my appreciation of the subtle art of piano tuning as carried out by a good piano tuner. I had often wondered if his 'art' contained an element of exaggerated mystique. Now I know how difficult it must be to do well. Certainly, electronic means seem overdue.

I wonder that Mr. Bedford did not mention the use of a digital frequency counter. Perhaps it would be too expensive? But would it not be feasible, at least in principle, after using Mr. Bedford's microphone, selective amplifier and squarer, simply to *read the answer*, instead of a 'sensed display' against a calibrated oscillator?

Using a frequency meter in this way would not necessarily limit the tuning to exact harmonic octaves (of even-tempered semi-tones) although it might be interesting to do this to hear what it sounds like. As Mr. Bedford says, exact harmonics can only be achieved by instrumentation. However, the 'octave stretch' which Mr. Bedford refers to could also be achieved by first tuning one octave precisely for correct read-out. Then the selective amplifier is switched to the second mode of the same octave played on the piano and the mode frequencies obtained, read and recorded. Then the next octave is played on the piano and tuned to the recorded values.

I wonder if Mr. Bedford would care to comment on the use of a digital frequency meter.

R. B. Newton

4 Breydon Walk, Furnace Green,
Crawley, Sussex.
9th October 1972

Clearly, 'laying the bearings' is much the most difficult part of piano tuning, and it seems to me that this could be greatly simplified using only a single crystal oscillator and a number of bistable counting and dividing circuits.

The musical tone would be picked up by a microphone, amplified and limited, and then divided by 128 in seven jitter-free divide-by-two stages. The output from the dividers would be used to start and stop a five-digit decade counter, counting the output of a 100 kHz crystal oscillator. A chart would be provided, giving the correct 'count' for each note of the bearing scale—for example, A49 (440 Hz) would register $128 \times 10^5/440$, or 29091, when correctly tuned. Alternatively, the instrument could incorporate a switch which, when set to 'A49' would put the counter in the state corresponding to 29091, and the acoustic signal would count it down to zero. In this case an auxiliary bistable would be needed to indicate whether the residual count was positive or negative in sign. The count itself would be quite rapid, and could be automatically repeated at a suitable interval—say, once every two seconds. Tuning the rest of the piano aurally would be relatively easy.

The principle described above has, of course, been used in a number of commercially available counters, and there could be problems over patents here, but technically such a device should be quite cheap and easy to make, using integrated circuits for the decade units and binary dividers, and I would have thought it would be quite easy to use—even by someone who was virtually tone-deaf!

H. Pursey

6 Branksome Way,
New Malden, Surrey.
12th October 1972

The tuning of a keyboard instrument is something which has been worrying me for some years, as I am an enthusiastic organ player and have at home a Hammond, which is not of the traditional type, but entirely solid-state employing dividers. However, having read various text books relating to both electronic instruments and keyboard instruments in general, I had been unable to obtain information to clarify the tuning.

Firstly, bearing in mind the relatively small error in tuning which occurs by assuming that all fifths have no beats and, therefore, a frequency ratio of 1–1·5, how small an error in tuning is, in fact, noticeable to the average person? Secondly, it is stated that the player of a musical instrument without a keyboard sounds a different pitch for 'G' sharp than he does for 'A' flat. If this is so, what, in fact, is meant by correct tuning from the musical point of view, assuming that one is not restricted by the twelve notes used on a keyboard instrument to form an octave?

A. Gwyn Jones

67 Witton Avenue,
Droitwich, Worcs.
18th October 1972

I am grateful for the various communications received in connexion with my paper on Electronic Tuning Aids, and beg leave to reply in detail only to the main point raised, namely 'why not a digital solution and read-out?'.

When I produced the Mk. III tuner some 12 years ago the possibility of a digital solution probably did not occur to me; with my c.r.t. background the course was obvious. Then when I reopened the attack in the last year on a cost-saving basis, it was the c.r.t. which had to go. The way was clearly open for a digital solution, but I dismissed this on grounds of cost. Only in the last few months, having seen the amazing outcrop of pocket digital computers, I realize that I could have been wrong in my cost 'estimates'; but this only if one could use the identical integrated circuits which are in quantity production for such purposes.

Yet, cost considerations apart, I remain unconvinced that a digital display is what is wanted. In my work in this field I have learnt

(a) that one cannot design a practical tuner on paper; field trials at the piano are absolutely essential. In particular

*'Electronic aids for the tuning of musical instruments', *The Radio and Electronic Engineer*, **42**, pp. S.131-5, September 1972.

(b) that any scheme which demands reference to a numerical table or schedule is unacceptable.

It follows that the digital read-out must be a difference frequency, something to be made null, in other words a beat-frequency as in aural or analogue-electronic tuning. One can formulate (as does Mr. Pursey) solutions to this problem, but the complication would seem somewhat vitiating. Even then it is questionable whether a numerical display of beat-frequency would be satisfactory. One likes to hear or see the beat-frequency move smoothly in response to movement of the wrest pin. With a numerical read-out one not only loses continuity of the display but there will surely be a proportion of false counts when the decrement of the string beats the counter. Again, I myself, and I suspect most tuners, like to hear as well as see what is going on. That a system can be operated by the totally tone-deaf does not seem to me to carry much appeal!

The two questions which Mr. Jones asks are related, and do not, I think, admit a simple reply. As to how small an error is noticeable, this is enormously dependent on the conditions of observation, in particular the pitch of the note concerned. But by and large one can say that one's sensitivity is vastly increased if the conditions allow the perception of *beats*. The best answer I can give to the combined questions, though not a very good one at that, is contained in my paper to the Royal Society of Arts in 1956 (Ref. 1 in the present paper).

L. H. Bedford

82A Hendon Lane,
London, N.3.
19th November 1972

Diplomas and Associateships

I would like to express my support for the timely warning from Mr. D. L. A. Smith in the September issue of *The Radio and Electronic Engineer*.* Without considering the details, I can summarize my support simply enough: I have worked on engineering projects of various magnitudes in England, the U.S.A., Sweden, France and Germany, and in every case the HND type of engineer is the one who was most valuable—or most sadly missed!

In the various countries in which I worked I tried to assess the respective merits and disadvantages of the systems of technical education available, and it would seem that the British system loses nothing by comparison with the others. The technical standards, the enthusiasm and involvement of the staff, and the watchful eye of the institutions; these are the corner-stones of good technical education, of which the British may be justly proud.

To 'europeanize' technical education just when the relatively weaker systems of intermediate technical education are looking to Britain for new ideas, can only sacrifice the one great asset which Britain brings into the Common Market—'know-how'. Britain must set the pace in unifying concepts and standards for technical education in Europe without sacrificing a sensible degree of diversity. Top level meetings are necessary, and much information must be gathered before such meetings commence. In the meantime the present system of technical education in Britain might be polished up, but under no circumstances should the diploma courses be sacrificed on the altar of uniformity.

Nigel J. Keen

Max-Planck-Institut für Radioastronomie,
Argelanderstrasse 3,
53 Bonn,
Federal Republic of Germany.
29th November 1972

* 'College diplomas and associateships in engineering, a matter of CEI policy', *The Radio and Electronic Engineer*, 42, No. 9, pp. S129–30, September 1972.

STANDARD FREQUENCY TRANSMISSIONS—December 1972

(Communication from the National Physical Laboratory)

December 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		December 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	−0·1	−0·2	+0·1	691	641·7	17	+0·1	+0·2	+0.2	736	656·4
2	−0·3	−0·1	+0·1	694	643·1	18	+0·3	+0·2	+0·1	733	654·5
3	−0·2	−0·2	+0·1	696	644·7	19	+0·3	+0·3	+0·2	730	651·3
4	−0·3	−0·2	+0·1	699	646·7	20	+0·3	+0·3	+0·2	727	648·2
5	0	−0·2	+0·1	699	648·9	21	+0·2	+0·2	+0·2	725	646·7
6	−0·2	−0·1	+0·1	701	650·1	22	−0·2	−0·1	+0·2	727	647·7
7	−0·1	−0·1	+0·1	702	650·8	23	0	−0·1	+0·2	727	648·7
8	−0·1	−0·1	+0·1	703	652·0	24	0	−0·1	+0·2	727	649·9
9	−0·3	−0·2	+0·1	706	653·5	25	−0·2	0	+0·2	729	650·1
10	−0·1	−0·2	+0·2	707	655·0	26	0	−0·2	+0·1	729	652·4
11	−0·2	−0·1	+0·2	709	656·3	27	−0·2	−0·1	+0·1	731	653·5
12	−0·1	−0·1	+0·1	710	657·7	28	−0·3	−0·2	+0·1	734	655·3
13	0	−0·1	+0·2	741	658·7	29	−0·1	−0·1	+0·2	735	655·8
14	+0·2	−0·1	+0·2	739	657·7	30	0	0	+0·2	735	656·0
15	+0.3	−0·2	+0·2	736	660·1	31	0	−0·1	+0·2	735	657·0
16	−0·1	+0·1	+0·2	733	658·2						

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

*** Relative to UTC Scale; (NTC_{NPL} − Station) = + 500 at 1500 UT 31st December 1968.**

† Relative to AT Scale; (AT_{NPL} − Station) = + 468·6 at 1500 UT 31st December 1968.

INSTITUTION OF ELECTRONIC AND RADIO ENGINEERS

Applicants for Election and Transfer

THE MEMBERSHIP COMMITTEE at its meetings on 29th November and 13th December 1972 recommended to the Council the election and transfer of 76 candidates to Corporate Membership of the Institution and the election and transfer of 29 candidates to Graduateship and Associateship. In accordance with Bye-law 21, the Council has directed that the names of the following candidates shall be published under the grade of membership to which election or transfer is proposed by the Council. Any communications from Corporate Members concerning these proposed elections must be addressed by letter to the Secretary within twenty-eight days after the publication of these details.

Meeting : 13th December 1972 (Membership Approval List No. 151)

GREAT BRITAIN AND IRELAND

CORPORATE MEMBERS

Transfer from Graduate to Member

BAVINGTON, Clive Robert. *Burton-on-Trent, Staffordshire.*
BRENT-JONES, John. *Andover, Hampshire.*
BURKE, Rodney John. *Reading, Berkshire.*
BURMAN, Arabinda. *Isleworth, Middlesex.*
CHAMPION, Peter Robert. *Ruislip, Middlesex.*
CHAPLIN, Robert John. *Biggleswade, Bedfordshire.*
DAS, Nirmal Chandra, M.Eng. *Hemel Hempstead, Hertfordshire.*
DAY, Malcolm Edward. *Maidstone, Kent.*
FITZGERALD, Alan George. *Luton, Bedfordshire.*
GILDER, Barry. *Hoddesdon, Hertfordshire.*
GLADWELL, John Martin. *Sandy, Bedfordshire.*
GROSSMITH, Barry Walter. *Harpenden, Hertfordshire.*
HATCHER, Ramon Alfred Frank. *Farnborough, Hampshire.*
HORNE, Kenneth Davidson. *Kirkcaldy, Fife.*
JONES, Michael Edward. *Royston, Hertfordshire.*
KING, John Edwin. *Harpenden, Hertfordshire.*
MANN, David. *Wantage, Berkshire.*
MARIS, Peter Ian. *St. Albans, Hertfordshire.*
MORRIS, John. *Stockport, Lancashire.*
PASHLEY, Brian Leslie. *Burgess Hill, Sussex.*
SAVAGE, David Robert William. *Broadway, Worcestershire.*
SHAH, Hasmukhlal. *London N11.*
SHARMAN, Kenneth Victor. *Harlington, Middlesex.*
SIMPSON, Keith Lewis. *Westbury-on-Trym, Bristol.*
TAYLOR, John Robert. *Wembley, Middlesex.*
WADE, Lance Edward, Major, REME. *Reading, Berkshire.*
WILSON, Christopher Geoffrey Charles. *Axbridge, Somerset.*
WRIGLEY, Richard Thomas. *Ware, Hertfordshire.*

Transfer from Associate to Member

LOCKYER, Walter Wallace. *Twickenham, Middlesex.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

GOLDER, Colin George. *Carshalton, Surrey.*

Direct Election to Graduate

BINDARI, Saad Abdo. *London SE9.*
BOUGHTWOOD, Martin Alan, B.Sc. *King's Lynn, Norfolk.*
CROWTHER, Nigel Joshua. *Moston, Manchester.*
CUTHBERT, Laurence Geoffrey, B.Sc.(Eng.). *Harlow, Essex.*
HIGNETT, David Ronald, Flt. Lt., RAF. *Bury St. Edmunds, Suffolk.*
OKOLI, Samuel Okechukwu. *London E15.*
SANDERS, Peter Philip. *London SW9.*
YOUNAN, Issam Saeed. *Reading, Berkshire.*

Direct Election to Associate

BATTEN, Mervyn. *Newbury, Berkshire.*
MACRAE, John. *Redruth, Cornwall.*
TURNER, Ronald. *Brierley Hill, Staffordshire.*

STUDENTS REGISTERED

BENNETT, John Nigel. *Ashton-under-Lyne, Lancashire.*
BROWN, Andrew Nigel. *Reading, Berkshire.*
HALL, Mervyn John. *Southampton, Hampshire.*
HUNTER, Graham. *Sheffield, Yorkshire*
KIRBY, Stephen Ashley. *Southampton, Hampshire.*
MANSUKHANI, Suresh Lokumal. *Thornton Heath, Surrey.*
MATTOCKS, David Ewart. *Maidstone, Kent.*
WELLS, John Robert. *Warrington, Lancashire.*

OVERSEAS

CORPORATE MEMBERS

Transfer from Graduate to Member

RODWELL, Walter Edward. *Grenada, West Indies.*

Transfer from Student to Member

THOMSEN, Roy Dudley. *Salisbury, Rhodesia.*

Direct Election to Member

't HART, Theodore. *Salisbury, Rhodesia.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

SOLLY, John Michael, B.Sc. *Sydney, Australia.*

Direct Electionn to Graduate

CHRISTOFOROU, Andreas Eracleous. *Limassol, Cyprus.*
EDELSTEIN, Aharon. *Haifa, Israel.*
EFSTATHIOU, Petros Demetriou. *Limassol, Cyprus.*
SIBBONS, Michael Edward, Lieutenant, REME, B.Sc.(Eng.). *BFPO 106.*

Direct Election to Associate

APPIAH, Robert Emmanuel. *Accra, Ghana.*
EZERENDU, Ikwuakolam Silvanus Ihekwereme. *Makurdi, Nigeria.*

STUDENTS REGISTERED

ABEYGUNAWARDENA, Kankanange Upajeewa Senaka. *Colombo.*
ARIYANI, Kumara Vidanalage Lalitha (Miss). *Dehiwala, Sri Lanka.*
CHAN, Cho Chun. *Selangor, West Malaysia.*
CHNG, Chew Lye. *Singapore.*
COOMARASWAMY, Brindha (Miss). *Colombo.*
DASSANAYAKA, Wadduwage Don Pusathie Soma (Miss). *Nugegoda, Sri Lanka.*
FONSEKA, Wanniachchige Dayaratne. *Colombo.*
KARUNATILLEKE, Weligamage Dayananda Chandrasiri Marthenesz. *Colombo.*
KHOGALI, Mohd. Osman. *Khartoum, Sudan.*
LEONG, Het Wah. *Kuala Lumpur, Malaysia.*
*LOW, Ah Kong, *Kuala Lumpur, Malaysia.*
PATHIRAJA, Pathirage Dona Padmini (Miss). *Bentota, Sri Lanka.*
RATNASEKERA, Gampahe Korallaya Chandrakumara. *Ratnapura, Sri Lanka.*
WIJEWARDENA, Weliwattage Indrani Anoma (Miss). *Pannipitiya, Sri Lanka.*
WIMAL, Liyanage Don Kithsiri. *Pita Kotte, Sri Lanka.*

**Reinstatement*

Meeting : 13th December 1972 (Membership Approval List No. 152)

GREAT BRITAIN AND IRELAND

CORPORATE MEMBERS

Transfer from Graduate to Member

ALDER, Christopher John. *Emsworth, Hampshire.*
ANDERSON, Geoffrey William. *Gillingham, Kent.*
ANDREWS, Alastair. *Grangetown, Cardiff.*
AUGHEY, Richard. *Lisburn, County Antrim.*
BERGER, Ernest John. *Margate, Kent.*
DAVIES, Hugh. *Altrincham, Cheshire.*
DIBBLE, Alan David. *Slough, Buckinghamshire.*
DRAKE, Antony John. *Enfield, Middlesex.*
ELMER, David Jeffrey. *Maidenhead, Berkshire.*
EMSLIE, Norman. *Ilkley, Yorks.*
HARRIS, Ian. *Devizes, Wiltshire.*
HEWITT, Patrick John. *Northwood, Middlesex.*
JEARY, David Andrew George. *Calne, Wiltshire.*
JOHNSON, Kenneth Walter. *London SW16.*
JONES, Elwyn John Brian. *Datchworth Green, Hertfordshire.*
KITTLE, David Victor. *London SW16.*
LE GOOD, Robert Keith. *Beckenham, Kent.*
LINCOLN, Raymond John. *Chalfont St. Peter, Buckinghamshire.*
MILLS, Eric Joseph. *Sittingbourne, Kent.*
PROCKTER, Adrian Claude. *London SE23.*
RICE, Stephen Alan. *Winsford, Cheshire.*
ROCHFORD, Michael James. *Keyingham, Yorkshire.*
RODGERS, Michael Daniel. *Whitehead, County Antrim.*
ROLLINS, Brian. *Wigan, Lancashire.*
SEYMOUR, William Lawrence. *Cookham, Berkshire.*
SMITH-HAMBLIN, Ronald Maurice, B.Sc. *Kempshott, Hampshire.*
TALGERI, Gurudutt Shripad. *London SE3.*
THOMPSON, Michael William. *Aylesbury, Buckinghamshire.*
TROUGHEAR, Alan. *Preston, Lancashire.*
WATERS, Christopher John Peter. *London NW9.*
WAYMARK, Richard David Burnham, B.Sc. *Enfield, Middlesex.*
WILLIAMS, Gerald. *Crowthorne, Berkshire.*
WISEMAN, Edward Elliott. *Warlingham, Surrey.*

Direct Election to Member

SINCLAIR, Clive Robert. *Edinburgh.*
TAPP, Brian William George. *Worcester Park, Surrey.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

COOK, Leslie Charles, B.Sc.(Tech.). *Malvern, Worcestershire.*

Direct Election to Graduate

MANKELL, Kenneth Clement. *Sheffield, Yorkshire.*
PLANT, Bernard Paul. *Ipswich, Suffolk.*
RODGER, Matthew Graham. *Milltimber, Aberdeenshire.*
WATERMAN, Bryan Charles. *Slough, Buckinghamshire.*

Direct Election to Associate

AZZOPARDI, Emanuele. *London N19.*
BROWN, Paul Anthony. *Leyland, Lancashire.*
OLLIVER, Tom, Flt. Lt., RAF., M.B.E. *Blandford Forum, Dorset.*
PARSONS, Colin Walter. *Witham, Essex.*

STUDENTS REGISTERED

CONTEH, Brimah. *Bolton, Lancashire.*
COYSTON, Andrew John. *Hertford.*
DILLIEN, Paul. *Barking, Essex.*
FIRMSTONE, Robin Peter. *London SW8.*
HOBBS, Michael Gregory. *Strood, Kent.*
HOLMES, Maurice Frank. *Rhondda, Glamorgan.*
HUDSON, Harry Thomas. *Hornchurch, Essex.*
JONES, Hugh Mansel. *Oakham, Rutland.*
LLOYD, Andrew Martin. *Timsbury, Somerset.*
MIGDAL, Richard. *Leicester.*
OH, Danny Siew Chu. *Guildford, Surrey.*
STEVENSON, Eric Graham. *Birstall, Leicestershire.*
SWINDELL, Paul. *Oadby, Leicestershire.*
THOMAS, Philip John. *Reading, Berkshire.*

OVERSEAS

CORPORATE MEMBERS

Transfer from Graduate to Member

GALEA, Joseph, B.Sc.(Eng.). *St. Paul's Bay, Malta.*
KELLY, John Henry. *Nairobi.*

LAU, Man Chung, B.Sc.(Eng.). *Hong Kong.*
ONYEGERE, Mpeji Clement. *Enugu, Nigeria.*
RAILTON, George Hugh, B.Sc. *Raumati, New Zealand.*
TUMOE, Sahr Raikes. *Freetown, Sierra Leone.*
VIJAY, Virendra. *Gulbarga, Mysore State.*

Transfer from Associate to Member

KENDALL-CARPENTER, Michael David Kendall. *Grand Cayman Island.*

Transfer from Student to Member

KANAGARAJAH, Kandiah. *Kajang, Malaysia.*

NON-CORPORATE MEMBERS

Direct Election to Graduate

MAHENDRAN, K. Nagalingham. *Kuala Lumpur, Malaysia.*

STUDENTS REGISTERED

CHENG, Choon Meng. *Singapore.*
CHEUNG, Shui Kay. *Osaka, Japan.*
DE SILVA, Weerawarnasuriya Patabendige Saddathissa. *Boossa, Sri Lanka.*
FERNANDO, Christopher Anthony Lester. *Battaramulla, Sri Lanka.*
GUNAWARDENA, Shanthakumara Dias. *Mt. Lavinia, Sri Lanka.*
KUMAR, Nanda L. *Bombay.*
KWEK, Heng. *Singapore.*
KWOK, Chi Pong. *Kowloon, Hong Kong.*
LIM, Kian Hock. *Singapore.*
PERERA, Athalage Sri Lal Mangala. *Nugegoda Sri Lanka.*
SITSABESAN, Sellappah. *Mt. Lavinia, Sri Lanka.*
TAY, Soo Meng. *Singapore.*
WAN, Tong Weng. *Singapore.*
YU, Kwok Chu Peter. *Kowloon, Hong Kong.*

Notice is hereby given that the elections and transfers shown on List 149 have now been confirmed by the Council.

CEI News

Engineers and Unions

At the meeting of the CEI Board on 26th October 1972, it was agreed that the following statement on CEI and the Trade Unions should be issued:

'The Council of Engineering Institutions is concerned with the welfare of the engineering profession generally and of Chartered Engineers in particular. It makes representation where necessary and desirable to Government and other bodies on behalf of the engineering community. It is not, under its Charter, a Trade Union. The Council recognizes that there are many circumstances where a Chartered Engineer may wish to be a member of some organization constituted for this purpose and is aware that considerable numbers of Chartered Engineers are in fact members of a trade union or professional association concerned with the negotiation of salaries and working conditions. It is realized that in many cases such membership is of long standing and has not in general given rise to conflict of interest with membership of the Institutions.

'The Council believes that the individual must retain the greatest freedom to join, if he wishes, a union of his choice which meets his needs, provided that a Chartered Engineer has at all times regard to his professional obligations and does not, in any action arising out of his membership of a union, contravene the CEI Code of Conduct and Rules of Professional Conduct of his own Institution'.

CEI Examination Regulations and Syllabus

The CEI Board have now approved new Examination Regulations and Syllabuses, to come into effect from the May 1974 Examination. Details of these are not yet available for general circulation, but advance information has been given to Polytechnic and Technical Colleges. Major changes in the Regulations include stricter conditions regarding withdrawal and absence from the examination, earlier closing dates and receipt by Institutions of applications for sponsorship, and a requirement that entrants for Part II of the examination must attend an approved course unless excused by their sponsoring Institution. As an offset to these stricter requirements, candidates will, from 1974, be permitted to take Part II in two sittings of three subjects each.

New Chairman and Vice-Chairman for CEI

Mr. T. A. L. Paton, C.M.G., F.R.S., C.Eng., has succeeded Sir Arnold Lindley, D.Sc., C.Eng., as Chairman of the Council of Engineering Institutions for 1973. The new Vice-Chairman of CEI is Major-General Sir Leonard Atkinson, K.B.E., C.Eng., Past President of the IERE.

Mr. Paton, who is a Past President of the Institution of Civil Engineers, is the Senior Partner of Sir Alexander Gibb and Partners, consulting engineers.

Sir Leonard has recently been appointed Chairman of the CEI Working Party on Student Recruitment which is to carry out an extensive examination of data drawn from such sources as Government departments, Industry Training Boards, UCCA, SRC, etc.

ERB Assess Qualifications of Technician Engineers in the Services

As part of its programme for the assessment of qualifications, other than those stated in its Bye-Laws (HNC or City & Guilds Full Technological Certificate for Technician Engineer; ONC or C. & G. Final Certificate for Engineering Technician) the Engineers' Registration Board recently set up a Committee to assess relevant Service qualifications.

The list of Service qualifications confirmed as meeting the requirements of the Technician Engineer Section—and holders of which are therefore eligible to apply for election to the Institution's grade of Associate—is as follows:

ARMY

Clerk of Works (Mechanical)	Foreman of Signals
Artificer Electrical Control	Artificer Radar
Artificer Telecoms	Artificer Avionics

(All subject to the completion of one year's experience at the appropriate level after qualification.)

ROYAL AIR FORCE

Maintenance Engineering (Electrical) Course (Eng 111)
Maintenance Engineering Course (Eng 112)
(Both subject to the completion of one year's experience at the appropriate level after qualification.)

The Services have gratefully accepted the offer of assistance from the Committee concerned in preparing the Defence Council Instruction promulgating its findings.

No Royal Navy qualifications in the appropriate categories were considered to meet the academic, training, and experience requirements for Technician Engineers. The reason for this is that they have in the past been tailored to the specific requirements of the Navy and it has been agreed that more attention will need to be paid in future to equivalence with civilian awards.

Forthcoming Institution Meetings

London Meetings

Wednesday, 7th February

JOINT IERE-IEE MEDICAL AND BIOLOGICAL ELECTRONICS GROUP

A Brief Review of Techniques in Foetal, Infant and Child Audiology

By Dr. R. J. Bench (*Royal Berkshire Hospital*)
IERE Lecture Room, 6 p.m. (Tea 5.30 p.m.)

Some differences in the physiological and psychological determinants of behaviour between the child and the adult will be used to explain the reasons for different audiological testing methods for the two age groups. This will be followed by a brief consideration of the merits of subjective versus objective (instrumental) methods in the child audiology clinic. Finally, apparatus will be described for hearing evaluation at various stages in childhood, from the foetus to the late pre-school child.

Tuesday, 13th February

JOINT IERE-IEE COMPONENTS AND CIRCUITS GROUP ONE-DAY COLLOQUIUM

TRANSISTORS—THE FIRST 25 YEARS

The Royal Society, 10 a.m.

Advance Registration necessary. Fee for members of the IERE and IEE £1.50, including a copy of the Colloquium Digest, morning and afternoon refreshments, and sandwich lunch (excluding beverages): the fee to non-members is £5. Application for registration to the IERE (see page 159 of this issue for further details).

Wednesday, 21st February

AEROSPACE, MARITIME AND MILITARY SYSTEMS GROUP

Meeting Cancelled

Wednesday, 28th February

AUTOMATION AND CONTROL SYSTEMS GROUP

Digital Phase Lock Loops

By K. R. Thrower (*Racal*) and P. Atkinson (*University of Reading*)
IERE Lecture Room, 6 p.m. (Tea 5.30 p.m.)

The digital phase-locked loop is widely used for frequency synthesis and control. The lecture is concerned with the design, performance, and practical applications of the loop. The shortcomings of the conventional Type 1 loop are discussed and it is shown how the Type 2 loop, which has integral control, overcomes these. Practical results are compared with computer predictions, where the sampling action of the phase detector is taken into account. In considering the applications of the digital phase-locked loops the problems of oscillator design, noise, jitter and frequency range are considered.

Wednesday, 7th March

COMMUNICATIONS GROUP COLLOQUIUM

OPTICAL COMMUNICATIONS

IERE Lecture Room, 2.30 p.m.
Advance Registration necessary

Thursday, 8th March

EDUCATION AND TRAINING GROUP

The Feedback Classroom

By K. Holling (*Chesterfield College of Technology*)
IERE Lecture Room, 6 p.m. (Tea 5.30 p.m.)

The Feedback Classroom is a group teaching/testing machine in which each student in a class is provided with a response unit so that responses to objective type questions can be made by the operation of switches. The responses are displayed to the teacher on a control console.

This lecture will describe the design and construction of a Feedback Classroom and its use for teaching and testing purposes. Examples will be shown of the objective test questions suitable for use at various levels ranging from craft to degree type courses.

Wednesday, 14th March

JOINT IERE-IEE MEDICAL AND BIOLOGICAL ELECTRONICS GROUP ONE-DAY COLLOQUIUM

IMAGE TECHNIQUES IN MEDICINE AND BIOLOGY

Morning Session: **Image Processing and Synthesis**

Afternoon Session: **Automatic Pattern Recognition**

IERE Lecture Room, 10 a.m.
Advance Registration necessary

Tuesday, 27th March

JOINT IERE-IEE MEDICAL AND BIOLOGICAL ELECTRONICS GROUP COLLOQUIUM

ARRHYTHMIA RECOGNITION AND DETECTION

IERE Lecture Room, 2.30 p.m.
Advance Registration necessary

Wednesday, 28th March

AEROSPACE, MARITIME AND MILITARY SYSTEMS GROUP COLLOQUIUM

SECONDARY RADAR IN MARITIME APPLICATIONS

IERE Lecture Room, 2.30 p.m.
Advance Registration necessary

Kent Section

Thursday, 1st March

Optoelectronics

By D. A. Bonham (*Texas Instruments*)
Medway College of Technology, Chatham, 7 p.m.

This lecture will cover a broad range of opto-electronics. Photosensors and emitters (both for visible light and infra-red), optical isolators and modules, and solid-state displays (both alpha-numeric and seven segment) will be discussed. The discussion will be applications orientated, showing possible uses for the devices. There will also be a few exhibits to demonstrate what can be achieved.

East Anglian Section

Wednesday, 7th February

JOINT MEETING WITH IEE

Feedforward: Yesterday's Techniques applied to Tomorrow's Amplifiers?

By Dr. T. J. Bennett (*Plessey*)
The Civic Centre, Chelmsford, 6.30 p.m. (Tea 6 p.m.)

The removal of unwanted distortion in a broadband amplifier using negative feedback depends on a comparison between the input and output being made at the input of the amplifier. With feedforward the input and output signals are compared at the output of the amplifier, and a second amplifier is used to give a correction signal. The basic principles of feedforward are discussed together with the advantages and disadvantages compared with other correction systems. Examples are given of particular applications from h.f. to X band. It is shown that feedforward can give performance which cannot be achieved with feedback, for example the correction of two identical amplifiers in a feedforward arrangement can give an improvement compared with a simple parallel connexion of the two amplifiers.

Thursday, 22nd February

JOINT MEETING WITH IEE

5 km Radio Telescopes

By Sir Martin Ryle (*Cavendish Laboratory*)
University Engineering Laboratories, Trumpington Street, Cambridge, 6.30 p.m. (Tea 6 p.m.)

Wednesday, 7th March

JOINT MEETING WITH IEE

Video Recording

By D. M. Bowd (*BBC*)
The Audio Visual Centre, University of East Anglia, Norwich, 7.30 p.m.

(PLEASE NOTE THIS IS A NEW VENUE AND TIME)

Wednesday, 14th March

Acoustic Surface Waves—the Prospects for Device Applications

By Professor E. A. Ash (*University College London*)
Department of Electrical Engineering, University of Essex, Wivenhoe Park, Colchester, Essex, 6.30 p.m. (Tea 6 p.m.)

The study of acoustic surface waves has now reached a stage of maturity. Whilst new phenomena and radically new concepts are still emerging, the current emphasis is on device development. In this lecture, both current applications, and the prospects for a wider usage of acoustic surface wave devices will be discussed.

Wednesday, 21st March

JOINT MEETING WITH IEE

The Transistor: its History and Consequences

By E. Wadham (*Mullard*)
Lecture Theatre 2, Civic College, Ipswich, 6.30 p.m. (Tea 6 p.m.)

The various eras of transistor development which have occurred over the past 25 years will be described with special emphasis on those aspects which have generated major advances. The contrast between the early germanium period which established the industry and the present situation of large scale integration involving computer aided design techniques will be highlighted by the showing of two Mullard films representative of these periods.

Thames Valley Section

Thursday, 15th February

Digital Communications in the Mobile Environment

By B. D. Parker (*Dollman Electronics*)
J. J. Thomson Physical Laboratory, University of Reading, Whiteknights Park, Reading, 7.30 p.m.

The transmission of data for the mobile radio service presents many problems to the design engineer. The wide variation of signal strength and ratio of signal to noise puts very exacting requirements on the signalling equipment. Many of the methods used successfully to transmit data over line or point to point radio links are unable to perform satisfactorily in this more difficult environment. The lecture will explain how some of the parameters of the data transmission equipment are influenced by the instability of this transmission path, and a logical method of signal encoding to overcome many of these problems will be developed.

Tuesday, 13th March

Modern Dynamic Measurement Techniques

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
J. J. Thomson Physical Laboratory, University of Reading, Whiteknights Park, Reading, 7.30 p.m.

Various methods of dynamic measurement widely used will be reviewed and the bases of their mechanization unified through an analysis of the cross-correlation process. Methods for predicting the noise rejection properties will be discussed, and an important difference between the serial and parallel modes of time domain measurement will be presented. The noise rejection capabilities of a range of available test equipments will be demonstrated.

North Western Section

Thursday, 15th February

Noise Reduction Techniques in Audio

By D. P. Robinson (*Dolby Laboratories*)
Lecture Theatre RG 7, Renold Building, UMIST, 6.15 p.m. (Tea 5.45 p.m.)

The lecture will consider the fundamental problems involved in the design of audio noise reduction systems, and show how the Dolby system technique overcomes these problems. The professional embodiment will be shown, and the further considerations involved in designing a consumer system will be discussed. The application to which the consumer system can be put to will be demonstrated. There will be many demonstrations including a simultaneous comparison of studio quality 15 in/s recordings with 1⅞ in/s cassette recordings of the same programme.

Wednesday, 21st February

Modern Dynamic Measurement Techniques

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
Room 137, Maxwell Buildings, Salford University, 2.30 p.m.
For Summary see Thames Valley Section, 13th March

Wednesday, 28th March

JOINT MEETING WITH IEE

Satellite Communication Systems

By Lt. Cdr. B. E. Collins, RN
Lecture Theatre RG 7, Renold Building, UMIST, 6.15 p.m. (Tea 5.45 p.m.)

Merseyside Section

Wednesday, 7th February

Self Organizing Control Systems

By Dr. D. W. Russell (*Liverpool Polytechnic*)
Department of Electrical Engineering and Electronics, Liverpool University, 7 p.m. (Tea 6.30 p.m.)

Wednesday, 7th March

Systems Control in the Electricity Supply Industry

By Dr. J. T. Boardman (*Liverpool University*)
Department of Electrical Engineering and Electronics, Liverpool University, 7 p.m. (Tea 6.30 p.m.)

North Eastern Section

Wednesday, 14th February

Electronics and Crime Prevention

A. T. Torlesse (*Plessey*)
Main Lecture Theatre, Ellison Building, Newcastle-upon-Tyne Polytechnic, 6 p.m. (Tea in Staff Refectory 5.30 p.m.)

Wednesday, 14th March

A Communication and Control System for Motorways

By E. H. Walker (*Department of the Environment*)
Main Lecture Theatre, Ellison Building, Newcastle-upon-Tyne Polytechnic, 6 p.m. (Tea in Staff Refectory 5.30 p.m.)

Yorkshire Section

Thursday, 15th February

JOINT MEETING WITH IEE

Induction Motor Speed Control by use of Permanent Magnetic Materials

By W. Shepherd (*University of Bradford*)
University of Leeds, 7 p.m.

Properties and applications of magnetically hard materials under cyclic excitation will be described.

Thursday, 22nd February

JOINT MEETING WITH IEE

Developments in Radio Telephone Communications

Speaker from Pye Telecommunications
YEB Offices, Hull, 6.30 p.m.

Thursday, 29th March

Radio Communication within the North Eastern Gas Board

By R. Grant (*NE Gas Board*)
NE Gas Board, New York Road, Leeds, 7.30 p.m.

East Midland Section

Tuesday, 13th February

25 Years with the Transistor

By Dr. K. J. Dean (*South East London Technical College*)
Room J.002, Edward Herbert Building, Loughborough University of Technology, 6.45 p.m. (Tea 6 p.m.)

The development of the transistor has been a strange mixture of careful research, happy chance, and the cut and thrust of commercial enterprise competing for high stakes. Against this patch-work an account will be given of the development of various types of transistor and integrated circuits over the last 25 years, together with some reasonable conjecture about the future.

Tuesday, 20th February

Modern Dynamic Measurement Techniques

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
Edward Herbert Building, Loughborough University of Technology, 7 p.m. (Tea 6.30 p.m.)
For summary see Thames Valley Section, 13th March

Tuesday, 20th March

Application of Digital Logic

By I. D. Brown (*Rolls Royce*) and S. L. Norman (*BPB Industries Instruments*)
Lecture Theatre A, Physics Block, Leicester University, 6.45 p.m. (Tea 6 p.m.)

Basic logic families are illustrated, with working models, culminating in transistor-transistor logic medium scale integration (TTL-MSI). This family is then used to produce a digital frequency meter having a 4½ digit readout and 20 MHz operating frequency.

West Midland Section

Monday, 19th February

Modern Dynamic Measurement Techniques

By Dr. J. A. Lamb and Dr. P. A. Payne (*UWIST*)
Department of Engineering Production, University of Birmingham, 6 p.m. (Tea 5.30 p.m.)
For summary see Thames Valley Section, 13th March

Monday, 26th March

Sonar and Underwater Acoustic Communications

By Dr. V. G. Welsby (*University of Birmingham*)
MEB Offices, Summer Lane, Birmingham, 6 p.m. (Tea 5.30 p.m.)

A review will be presented of modern techniques based on the use of sound waves in the sea and in lakes, rivers, etc. Systems for diver communication and navigation are described. High resolution sonars, sometimes using focused acoustic arrays, have uses which range from the study of the behaviour of fish shoals to aiding police searches in muddy canals. Acoustic telemetry is used to control submersible vehicles and to channel collected information back to the surface. Acoustic waves are used to count migrating fish in rivers.

South Midland Section

Tuesday, 13th February

JOINT MEETING WITH IEE

How High is Hi-Fi?

By D. Aldous (*Hi-Fi News and Record Review*)
BBC Club, Evesham, 7.30 p.m.

Tuesday, 13th March

(PLEASE NOTE CHANGE OF DATE)

Telecommunications in the Year 2000

By A. G. Hare (*PO Research Station, Dollis Hill*)

Abbey Hotel, Malvern, 7.30 p.m.

South Western Section

Thursday, 1st February

JOINT MEETING WITH IEE

Marine Satellite Communication Systems

By Dr. W. P. Williams (*Marconi Marine*)
Plymouth Polytechnic, 7 p.m. (Tea 6.30 p.m.)

The use of satellites for merchant marine communications is still very much in the experimental stage, but work is progressing rapidly and already decisions have been made on the frequency bands to be used and a certain amount of work has already been done to determine the broad parameters of the final system. This paper reviews the work carried out to date and gives some assessment of the international agreements likely in the near future which will affect the form which the final service will take. The paper then attempts some assessment of the likely forms which the ship-borne and satellite equipment may take. Finally, there is an attempt to assess likely future programmes and the dates on which the first full commercial service is likely to be available.

Wednesday, 21st February

JOINT MEETING WITH IEE AND RAES

Concorde Flight Test Data Recording Systems and their Use

By R. McKinlay *et al.* (*RAF Locking*)
RAF Locking, 6 p.m.

The team presentation will be organized by Mr. R. McKinlay and based on experience gained with 'onboard' data systems in *Concorde*, but may also cover aspects of telemetered recording, as well as system design, data analysis and certification evidence.

Wednesday, 7th March

JOINT MEETING WITH IEE

Sound in Syncs

By Dr. C. J. Dalton (*BBC*)
Bath University, Room 2E.3.1, 7 p.m. (Tea 6.30 p.m.)

'Sound-in-syncs' is a system developed by the BBC whereby pulse code modulation techniques are used to combine the sound and picture components of a television programme in a form suitable for distribution over a single vision circuit.

Wednesday, 14th March

Modern Dynamic Measurement Techniques

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
Bath University, Room 2E.3.1, 7 p.m. (Tea 6.30 p.m.).
For summary see Thames Valley Section, 13th March

Southern Section

Friday, 9th February

Acoustic Surface Wave Devices and Applications

By Dr. J. Heighway (*Allen Clark Research Centre, Plessey*)
Newport IOW Technical College, 7 p.m.

Acoustic surface wave devices are finding an increasing number of applications in signal processing systems. These devices utilize the flow acoustic velocity to produce miniature delay lines and filters. In addition the surface wave nature of the devices makes more sophisticated signal processing possible. Applications range from pulse compression filters through bandpass filters to oscillators. The lecture will concentrate on the devices and on their present and future applications.

Wednesday, 14th February

JOINT MEETING WITH IEE

Design of British Scientific Satellite

By D. J. McLauchlin (*Marconi*)
Portsmouth Polytechnic, 6.30 p.m.

Wednesday, 28th February

Port of Southampton Signal and Radar Station

By D. J. Doughty (*B.T. Docks Board*), J. C. Gunner (*Decca Radar*) and J. R. Laver (*Marconi*)
Geography Lecture Room G1, University of Southampton, 6.30 p.m.

The port of Southampton has a unique new port navigational system integrating 11-channel v.h.f. and computer-assisted radar installations at the new Port Operation Centre operated by the British Transport Docks Board. Radio and radar coverage of the port area and Solent to the Nab Tower will be described. Mr. Doughty will discuss the marine requirements and the British Transport Board's responsibilities, etc.; Mr. Gunner will speak on the design and provision of the radar system; and Mr. Laver will describe the design and installation of the communication system.

Saturday, 3rd March

Technical Visit to Southampton Docks Radar Installation, 10.15 a.m.

Apply for tickets at the lecture on 28th February or to Mr. R. S. Broom, Department of Electronics, University of Southampton (Southampton 59122 Ext. 540)

Tuesday, 6th March

Advances in MOS Device Technology

By Dr. D. R. Lamb (*Southampton University*)
Brighton Technical College, 6.30 p.m.

At present the microelectronics industry has available a number of different fabrication techniques for the manufacture of m.o.s. devices and circuits. Thus in optimizing the design of a device or circuit an engineer should be aware of these different technologies and their respective advantages and disadvantages. In this talk these technologies will be described and compared, and their use in both conventional and more novel applications will be discussed. This will include the basic aluminium gate m.o.s.t., m.n.o.s. devices, autoregistered m.o.s.t.s—both metal gate and silicon gate, and charge coupled devices.

Monday, 12th March

Modern Dynamic Measurement Techniques

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
HMS Collingwood, Fareham, 6.30 p.m.
For summary see Thames Valley Section, 13th March

Wednesday, 14th March

JOINT MEETING WITH IEE

Communication U.H.F. Modules

By P. Tunbridge (*Mullard*)
Lanchester Theatre, University of Southampton, 6.30 p.m. (Tea in Senior Common Room, 5.45 p.m.)

Wednesday, 21st March

JOINT MEETING WITH IEE

Application of Control to Artificial Limbs

By Professor J. M. Nightingale (*University of Southampton*)
Lanchester Theatre, University of Southampton, 6.30 p.m. (Tea in Senior Common Room, 5.45 p.m.)

Thursday, 29th March

ANNUAL GENERAL MEETING (6.30 p.m.)

Followed by **'Aspects of Stereo Broadcasting'**

By J. H. Brookes (*BBC*)
Farnborough Technical College (7 p.m.)

South Wales Section

Wednesday, 14th February

ANNUAL GENERAL MEETING (6 p.m.)

Followed by **A Short-Hop Radio-Relay System at 20 GHz**

By R. R. Walker (*PO Radio Research, Castleton*)
UWIST, Cardiff, 6.30 p.m. (Tea 5.30 p.m.)

The feasibility of radio-relay systems at frequencies around 20 GHz is being evaluated by the Post Office. An experimental system using digital modulation at 120 Mbaud is described. The repeaters use solid-state equipment mounted on steel poles up to 30 m in height and spaced between 5 km and 10 km apart.

Monday, 19th March

JOINT MEETING WITH IEE

Modern Measurement Techniques in Control Engineering

By Dr. J. D. Lamb and Dr. P. A. Payne (*UWIST*)
UWIST, Cardiff, 6.30 p.m. (Tea 5.30 p.m.–6.30 p.m.)
For summary see Thames Valley Section, 13th March

Scottish Section

Thursday, 8th February

CEI LECTURE

The Work of the Design Council

By Hugh Conway and Sir Paul Reilly
University of Strathclyde, John Anderson Building, Taylor Street, Glasgow
6.15 p.m. Tickets *not* required

The speakers, who are Deputy Chairman and Director respectively of the Design Council, will deal with the activities of the Council, with particular reference to its expanding work in the field of engineering design. They will call for better communications between designer and manager. Mr. Geoffrey Constable, who heads the Design Council team working on engineering design, will participate in the discussion session.

Monday, 12th March

JOINT MEETING WITH IEE

The Industrial Uses of Lasers

By Dr. I. J. Spalding (*Culham Laboratories*)
Room 406, James Weir Building, University of Strathclyde, Glasgow, 6 p.m.

Repeated on
Tuesday, 13th February

South of Scotland Electricity Board Showrooms, 130 George Street, Edinburgh, 6 p.m.

Tuesday, 6th March

JOINT MEETING WITH IEE

Seismological Measurements

By Dr. P. L. Willmore (*Siesmological Centre, Edinburgh*)
Robert Gordon's Institute of Technology, St. Andrews Street, Aberdeen, 7.30 p.m.

Repeated on
Wednesday, 7th March

Napier College of Science and Technology, Colinton Road, Edinburgh, 7 p.m.
and on
Thursday, 8th March

Glasgow College of Technology, North Hanover Street, Glasgow, 7 p.m.

Wednesday, 21st March

CEI LECTURE SPONSORED BY IERE AND IEE

Fifty Years of Broadcasting

By J. Redmond (*BBC*)
The Boyd Orr Building, University of Glasgow, 7 p.m.

Northern Ireland Section

Tuesday, 13th February
(PLEASE NOTE CHANGE OF DATE)

A Discussion Forum—**Reliability in Electronics—Fact or Fiction?**

Cregagh Technical College, 7 p.m.

Tuesday, 27th March

The Eric Megaw Memorial Award

Four student papers will be presented
Ashby Institute, Stranmillis Road, Belfast, 6.30 p.m.

Forthcoming Conferences and Exhibitions

The Physics Exhibition 1973

For the first time, the Institution will have a stand at the Physics Exhibition, to be held at Earls Court, London, from Monday, 9th April to Friday, 13th April. Institution publications will be on display and information on membership, etc. will be given. LABEX International 1973 will take place concurrently in the same building.

Admission to the Exhibition will be by season ticket, obtainable free of charge from the Institute of Physics, 47 Belgrave Square, London SW1 8[illegible]; members of the IERE may obtain tickets from the Institution, 9 Bedford Square, London WC1B 3RG. In both instances, a stamped addressed envelope with requests for tickets would be helpful.

Teaching of Electronic Engineering in Degree Courses

The above Conference is being organized at the University of Hull by the Department of Electronic Engineering, together with representatives of the British Computer Society, IEE and IERE and will take place from 11th to 13th April 1973.

Some 30 papers will be presented during eight sessions and will include contributions from both industrialists and academics. A Conference fee of £5 will be charged and includes a copy of the Proceedings. Accommodation and meals, including a Conference Dinner, will cost £15. Further details and registration forms can be obtained from Dr. F. W. Stephenson, Department of Electronic Engineering, The University, Hull HU6 7RX.

Conference on Noise and Vibration Control

This third Conference follows those held in Cardiff (1971) and Bath (1972). It is jointly sponsored by the University of Bath, the Institution of Radio and Electronic Engineers, the Society of Environmental Engineers and the University of Wales Institute of Science and Technology and will take place at UWIST, Cardiff on 16th–18th April 1973. The ground covered by the Conference has been enlarged. It will seek to increase awareness of noise and vibration problems and means of assessment and solution in industry but will also deal with some problems of particular interest to local authorities.

The principles of noise and vibration control as applied to industrial problems will be covered by means of case studies. Aspects of the law on noise and vibration, criteria and standards laid down by government and national bodies will be considered.

The third day will be devoted to noise in buildings with emphasis on the practical aspects of its control.

Further details are available from: The Conference Secretary (Mrs. J. Jackson), UWIST, King Edward VII Avenue, Cathays Park, Cardiff CF1 3NU. (Telephone: 0222-42522, ext. 211.)

Microwave 73 International Conference and Exhibition

The IERE are to be associated with the international conference and exhibition for the microwave industry, Microwave 73, which is to be held in Brighton from June 19th to 21st. It will be opened by H.R.H. the Duke of Kent, a member of the National Electronics Council and a Companion of the IERE.

The conference will be an applications-oriented event concentrating on the research, design, development, techniques and applications of microwave components, devices, instruments, sub-systems and systems. The European Papers Committee has been joined by an IERE representative, Dr. J. R. James of the Royal Military College of Science.

Special reduced registration fees for the conference will be offered to IERE members namely £25 which includes a bound pre-printed copy of the proceedings. Further information from Conference Department, IERE, 8–9 Bedford Square, London WC1B 3RG.