THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

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 28 JULY 1964 NUMBER 1

The Fifth Clerk Maxwell Memorial Lecture

by

Sir Gordon Radley, K.C.B., C.B.E., Ph.D.(Eng.) (*Honorary Member*)†

Presented at a meeting of the Institution in London on 16th May 1964.

Introduction

I would start by expressing my appreciation of the honour which the Institution of Electronic and Radio Engineers has done me by its invitation to deliver the Clerk Maxwell Memorial Lecture. To be associated in a list of Memorial Lecturers with such names as Sir John Cockcroft, Sir Lawrence Bragg and Dr. Vladimir Zworykin is in itself a distinction, but all of us who have attempted—or who are attempting—to give this Lecture must gain a little reflected glory from the name of the man it commemorates.

The choice of subject has been a difficult one. One of the exciting characteristics of contemporary electronics is that new avenues of rapid technological advance are being opened up at increasingly frequent intervals by discoveries in electrical physics. Also of importance is the way in which electronics has infiltrated into heavy electrical engineering and many other older technologies. All this indicates the very wide field of interest covered by your Institution. It is one that is far too wide to deal with in a single lecture, so that a choice of something much narrower becomes necessary.

I was tempted to deal with world communications, that being an activity with which I have had some connection during my working life. Trans-oceanic communication cables of high-traffic capacity, equipped with submerged electronic repeaters, have made their contribution to the development of electronic technology; in particular, they have encouraged the production of components of very high reliability and long life. They also have the commercial attraction of providing comparatively cheap high-quality world telephone circuits. But because of my vested interest in these developments it might be concluded that I was not giving proper attention to the possibilities of satellites for a new system of world communication; although I must comment here that, quite apart from the commercial value of satellite systems, the technological 'know-how' acquired in the construction of such advanced systems must add to technological ability in other electronic fields. However, I have decided to talk generally about the development of computers and industrial automation. This is a field in which I have become greatly interested. It is also one of very great significance for the future.

When Mr. Clifford, your Secretary, wrote to me conveying the Institution's invitation to give this address he said he hoped that I would point to future developments. I will endeavour to do so as an engineer, for this is the discipline in which I have been taught to think. But as I do so you may be tempted to think that I have overlooked my other task—that of honouring James Clerk Maxwell. This is because in talking about computing processes we shall be following Nature in thinking digitally—in terms of discontinuities. He, on the other hand, showed that the propagation of electromagnetic waves resulted from the inter-relation of continuously varying electric and magnetic fields. Nevertheless, although some of Maxwell's physical explanations have fallen out of science—as the scaffolding is abandoned when the building goes up—his conclusions remain as the basis of a great deal of contemporary engineering achievement. They are encountered in the field of data-handling.

† Chairman of the Marconi Company Ltd., London.

As I look ahead I shall endeavour to show that the embodiment of microminiature techniques in computers of the future will lead to machines that are more compact and therefore able to operate at much faster rates. Because micro-electronics can cut out many wire interconnections they should mean more reliable and eventually cheaper machines. All this should lead to more ambitious computer projects, and the use of computers in locations which cannot accommodate the large machines of 1964.

Electronic Data-Processing as a Major Advance

In a recent address to the Electronics Division of the Institution of Electrical Engineers on "Electronics—the expanding frontier", Dr. R. C. G. Williams referred to advances that have resulted in 'million-to-one gains' in our ability to do things. Certain of these in history come readily to mind—the invention of printing, the steam engine and the industrial revolution, Faraday's experiments on electromagnetic induction leading to the practical generation and use of electrical power, the first flight by a heavier-than-air machine, the release of energy from the atomic nucleus, and so on. Quantitatively electronic data processing constitutes a gain of this order; in one second a present-generation machine for business office applications can add or subtract a million times, multiply 70 000 times, or sort 6000 items from random order. Obviously, the availability of equipment which will do this must result in great changes in our commercial, industrial and social life. Such equipment is an essential aid to defence against modern forms of enemy attack.

The pace of progress in data handling has so far been set by such defence requirements as detection and course prediction of enemy missiles and tracking and guidance of one's own. Space developments require much data handling for the guidance of satellites. For these kinds of problems speed of calculation and subsequent operation are of the utmost importance, and sometimes they require the operations to be accomplished in part by equipment occupying the minimum volume. Work in the United States has been heavily sponsored by the Government. By its encouragement of micro-electronics the U.S. Government started off a leap-frog movement from conventional methods to something that stretches the inventive and investigational abilities of the physicist, the chemist and the electronics engineer. American programmes on machines are supported by equally impressive programmes of research into fields of pure physics—on cryotrons and related super-conductive devices, on magnetic films, on tunnel diodes, and so on. Many of the possibilities involve the use of low-temperature techniques. But these are examples only and do not exhaust the list of new devices which may be used in future computing circuits and give rise to a generation of machines of higher speed and greater capacity than any machines at present available. While new generations of computers may be developed primarily to meet military needs, they will also claim a substantial share of the civil market of the future. In the United Kingdom the Department of Scientific and Industrial Research has initiated a programme involving cooperation between Government laboratories and industry. Of necessity it is on a much smaller scale than contemporary American work.

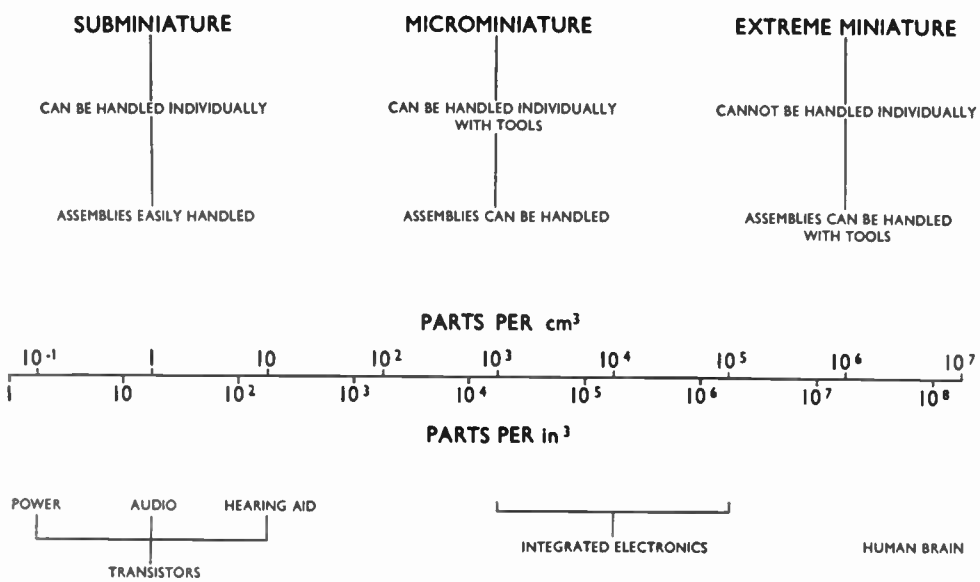


Fig. 1. Packing density of electronic components.

Micro-electronics

The searches for higher speed, greater reliability and smaller size of overall equipment have all converged towards the use of microscopic components placed close together until the point is reached when inter-component wiring becomes unrecognizable as such.

It is worth recalling that for over twenty years considerable efforts have been devoted to increasing the component density of electronic equipment. The traditional wired circuit of the 1940s possessed a component density of the order of 0·1 per cubic inch. In the early 1950s, using printed circuits, this figure had risen by ten times. The replacement of valves by transistors and the use of sub-miniature components enabled a component density of between 30 and 50 per in^3 to be attained. Using the term 'micro-electronics' as comprehensive of the new thin-film techniques and solid-state microcircuitry, micro-electronics has made packing densities of 1000 to 10 000 components per in^3 theoretically possible. Even this does not approach the packing density of memory elements in the human brain.

Naturally, with so many sciences and technologies involved, there have been many different approaches to the microminiaturization of circuits and components. But in the general stream of progress, and stopping short of possible low-temperature and other future devices, two main techniques may be recognized, although their identification is by no means indicative of the whole picture, even as it exists today. As usually defined, 'thin-film' microcircuits use passive substrates and thin-film passive components of metal or metal-oxide. Active elements are added separately. 'Solid-state' microcircuits are those using semiconductor, usually silicon, substrates with diffused active elements, passive elements being either diffused or thin-film.

There seems to be a fair consistency of opinion in the United States that the annual turnover of the microcircuit business will reach 750 million dollars by 1970. Of this 90% will be in respect of units employing basically semiconductor techniques. It is interesting to note that one survey predicts that 35% of the output will be absorbed by the industrial equipment market (as distinct from consumer goods) and 45% by military uses.

A Contemporary Microcircuit Technique

Time will not permit any detailed discussion of the various techniques during a comparatively short dissertation. I will instead concentrate on what is a very useful concept at this time and one that has been very successfully developed to the point of embodiment of microminiature units in equipment. I am conscious that in doing this I am giving a hostage to fortune because what is contemporary this evening may well be out of date by the time this lecture is printed—so fast is the art moving!

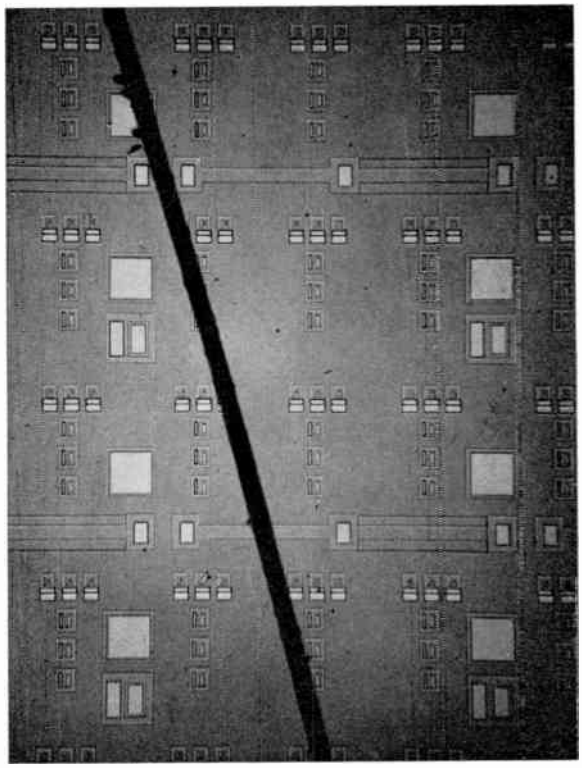


Fig. 2. Matrix of high-frequency transistors with human hair (0·003 inch diameter) superimposed.

The technique which I am going to describe owes its origin to the introduction of the silicon planar process for transistor production. It has led over a period of four years to the fabrication of truly microminiature units, usually between 50 and 75 thousandths of an inch square, although some of them may be much smaller. They are made in matrices of hundreds of units on a single slice of silicon.

In this technique high-quality silicon dioxide is grown thermally on an n-type silicon substrate and used as a diffusant mask against boron and phosphorus, which are employed respectively as p- and n-type additives in accordance with normal transistor manufacturing processes. 'Windows' are opened in the oxide by photolithographic and etching processes, to permit ingress of the additives into selected areas of the substrate by high-temperature diffusion. It is arranged that the windows are closed at the termination of each diffusion, being freshly opened for the subsequent processes. By this means the base and emitter of a transistor and alloyed aluminium contact

areas are all prepared. Diffusion depths are commonly around 3 microns, and the differential penetration of phosphorus and boron may be about one-half micron in a high-frequency device. The alloy contact areas can be as small as 0·5 mil square, and the photolithographic processes used have been refined until a fine sub-structure of 0·25 mil is permissible in normal device design. Sequential registration tolerances for the series of photolithographic operations are required to be better than one-tenth of a mil and this figure is now a common working limit.

The attraction of this method of fabrication lies in its general application to other circuit elements. These can be regarded as part-finished transistor structures which can be made simultaneously with the transistors on the same slice, if desired. Diodes and resistors are produced by a single diffusion. Capacitors may be produced either by diffusion or preferably by utilizing the silicon dioxide masking layer as a dielectric with an aluminium overlay. The degree of microminiaturization has gone beyond the scale factor where it could be applicable to thin-film inductors.

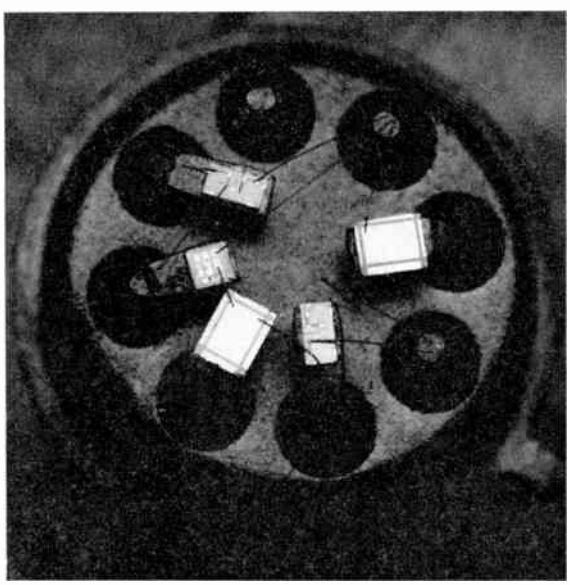


Fig. 3. 'Multi-chip' construction used for encapsulated logic circuit.

For tight tolerances, high-speed or high-frequency circuits a 'multiple chip construction' is used. This starts with the selection of 'chips' of semiconductor, containing pre-tested components which are then mounted either on a ceramic disc with preformed gold mounting areas or on the pins of a suitable header. Interconnection between chips, or chips and pins, is effected by thermo-compression bonding of gold or aluminium wires. Individual chips will contain groups of components such as resistor chains or common-collector transistors, and the choice of fabrication technique and substrate is open to maximize yield for each type of element. This construction has resulted in logic stages with under 5 nanoseconds propagation delay and has also been used in linear amplifier circuits at 75 Mc/s.

A high order of reliability is claimed for the micro-circuits re-integrated on the 'multi-chip' technique within a header. All p–n junctions associated with active and passive devices are stabilized and protected by a thermally-grown oxide and the parameters of devices formed below the surface of the substrate are determined by high-temperature diffusant concentrations. Both of these features lead to good long-term stability. The thermocompression bonds, which at first assessment might be thought a weak point in the construction, have been shown to withstand accelerations in excess of 20 000 *g* under centrifugal test.

Fully Integrated Circuits

A natural progression from the point we have reached is to contain two or more circuit components —say, a number of transistors, or a combination of transistor and resistor—on a single chip, thus eliminating the wire connections between these particular components. This can easily be arranged by appropriate masking during the successive processings. It leads to the fabrication of fully-integrated circuits—that is, those in which all the components comprising a simple circuit are contained within a single piece of semiconductor. In this way it is, for example, possible to construct a simple gate circuit on a silicon chip 0·05 in square by six successive processings. These circuit units may be produced as elements of a matrix on a single slice of semiconductor. There is still the proviso that we have not yet learned how to include inductance in this way.

With the fully-integrated circuit there are, of course, no internal wire connections, and this gives promise of a further increase in reliability. But, at the present point in time, the advantages of the 'multi-chip' technique, as compared with the fully-integrated circuit, are:

(i) the production yield to specification is much higher;

(ii) tolerances can be more readily set for the individual components as well as for the overall performance of the re-integrated circuit;

(iii) the technique is more flexible and allows experimental circuits to be made for the equipment design engineers at much less cost than that of fully-integrated circuits.

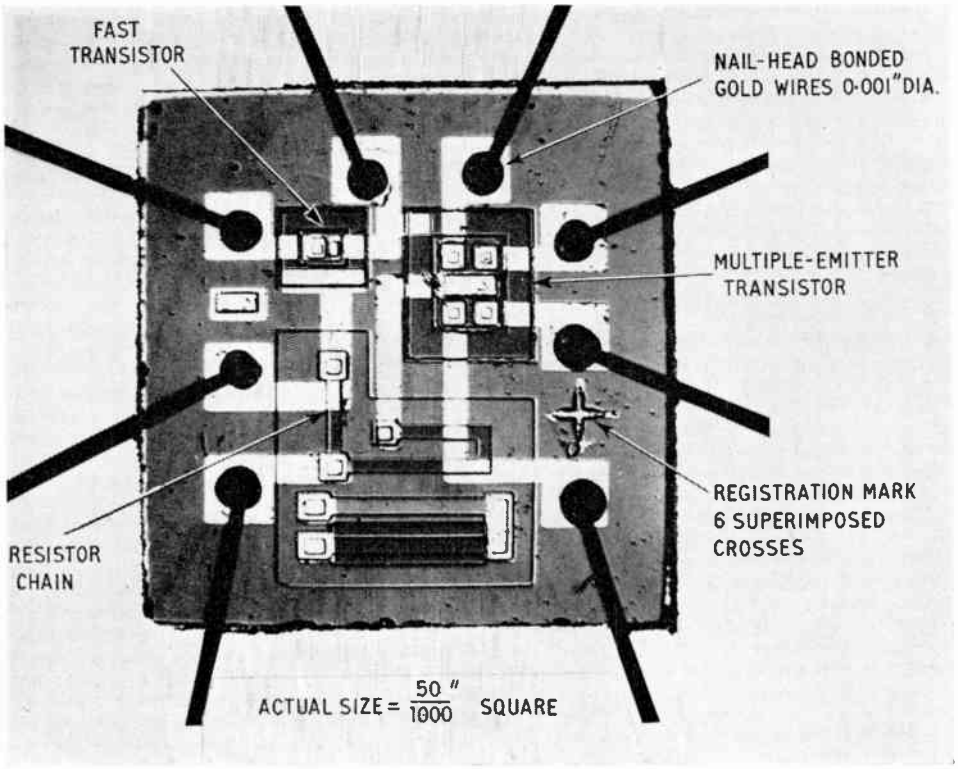


Fig. 4. Multiple-emitter gate on single silicon chip.

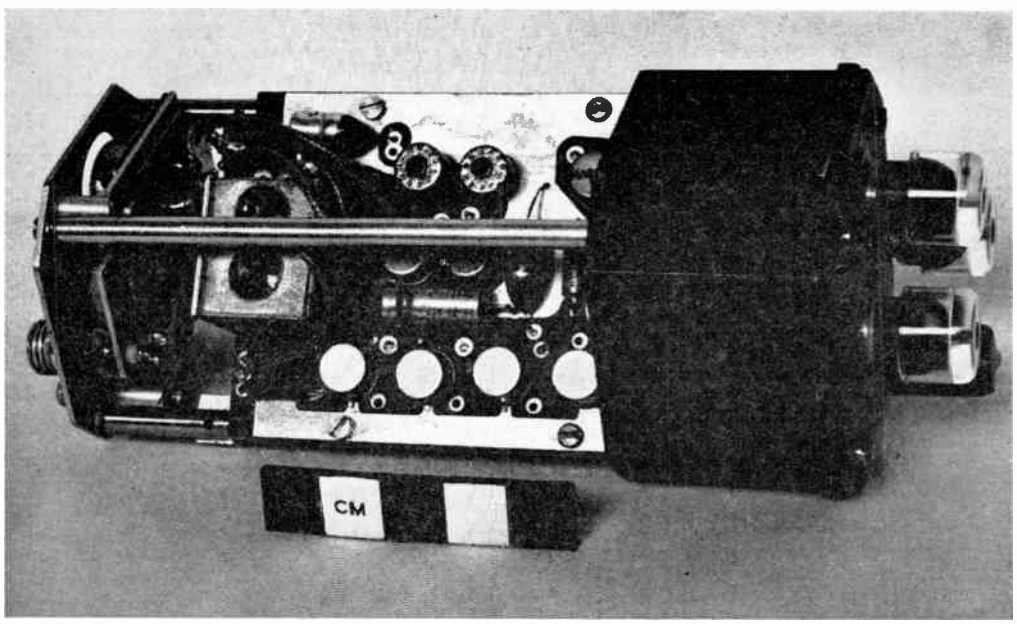


Fig. 5. 75-Mc/s receiver constructed with multiple-chip micro-electronics for use in aircraft instrument landing system.

'Thin-film' Microcircuits

The semiconductor microcircuit technique which I have described is tending to overshadow its older rival—the thin-film microcircuit on a glass substrate—to which semiconductor active elements had to be added separately. But with the replacement of glass by a silicon substrate, insulated on the surface by a silicon dioxide masking layer, the two techniques can now be said to have converged. It is believed that hybridization of thin-film and semiconductor microcircuits, using a common semiconductor substrate, will be of increasing importance, and, since mask-making, assembly and bonding are already common technology, it represents a natural evolution.

Storage of Information

There are two important things about a computer; they are first its ability to do arithmetic and second its capacity for remembering information fed to it, together with the rapidity with which we can obtain access to that information. With regard to its memory, the electronic engineer has still a long way to go to catch up with the human brain in respect of the number of bits of information stored in a given volume.

The first real break through in the compact electronic storage of information came in 1945 when Professor F. C. Williams invented a method of storing information on, and reading it from, the face of a cathode-ray tube. Since then many methods have been tried, but the use of threaded magnetic cores has become almost universal. Each core constitutes, in effect, a 'yes' or 'no' point in a matrix.

New developments in computor memories indicate that progress is being made in their fabrication compatible with the development of integrated circuits for the arithmetic function. Lack of this capability has been a stumbling block in progression to the microelectronic computer of the future. Memory research is aimed at increasing speed and capacity—a goal which can be met only by reducing the size of the memory elements. Magnetic and super-conductive techniques are receiving a lot of attention. While magnetic techniques look promising for high-speed random-access memories with capacities up to around 10^7 bits, super-conductive thin films obtained in low-temperature units may be the answer for large capacity memories of the order of 10^9 bits. But low-temperature stores have their obvious disadvantages which may weigh heavily against their being a success commercially. Very large stores based on any other purely electronic or electro-optical scheme do not seem in sight (but one can be very wrong indeed) and so ferrite core stores seem to be the correct choice, at least for the next generation of computers. A good deal has been published recently about improvements in fabricating them.

Computers of the Future

1. *Engineering Form.* The next question is: "What kind of computer is likely to be built using the new generation of components?" In attempting to answer this question I may fail to distinguish between machines that are just 'around the corner' and those which are much farther away in time. This is inevitable. Every engineer will appreciate the many technical snags that are encountered in the conversion of a promising laboratory concept into equipment coming off the production line. Both the time and also the cost involved in this are often under-estimated.

Starting with the logic circuits, we quickly come up against an apparent paradox. We have seen a move towards the microminiaturization of components and circuits and, within the limits imposed by ease of handling, optical resolution and bonding techniques, the cost should come down as the size becomes smaller. But, since the smaller the device the higher its frequency capability, we can say that, within limits, low cost and high performance will tend to go together.

Here we begin to come up against the question of the speed with which a bit of information can be transferred from one point to another. This kind of problem has, of course, for long been in the minds of communication engineers, particularly those responsible for providing long-distance communications. The time taken for a radio signal to travel half-way round the world to Australia is about 60 milliseconds; if it should travel there via two communication satellites revolving in synchronous orbits both the distance and the time delay will be greatly increased—the latter to nearly half a second—and communication engineers argue whether this is serious in telephone conversation. Also, should the signal, instead of being transmitted freely through space, be sent along the copper conductor of a submarine cable so that the electromagnetic wave is surrounded by dielectric and within the submerged electronic repeaters has to traverse filter networks, the speed of travel is reduced and the transmission time for the 14 000 mile journey from London to Sydney, routed via the CANTAT and COMPAC cables, is nearly 150 milliseconds.

But these transfers involve great distances and within a computer we are not concerned with distances of more than a few feet. On the other hand, a nanosecond has become a significant time-interval and in a nanosecond Clerk Maxwell's equations show that, even in free space, a signal will not travel farther than about 1 ft. Therefore, in order to utilize the

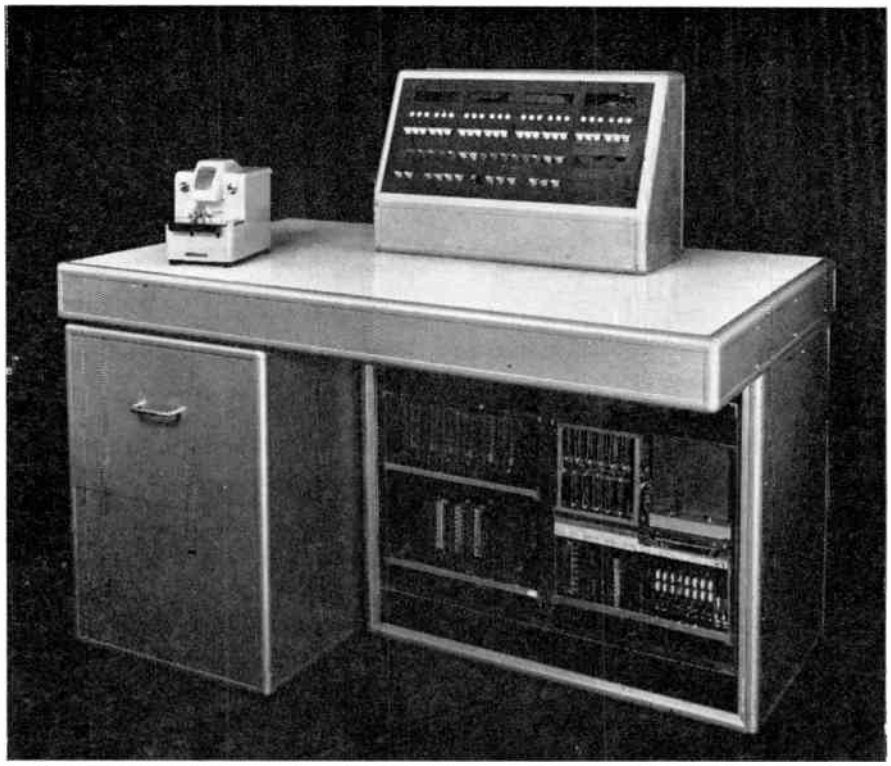


Fig. 6 (*a*). Small high-speed computer using silicon micro-logic devices described in the paper.

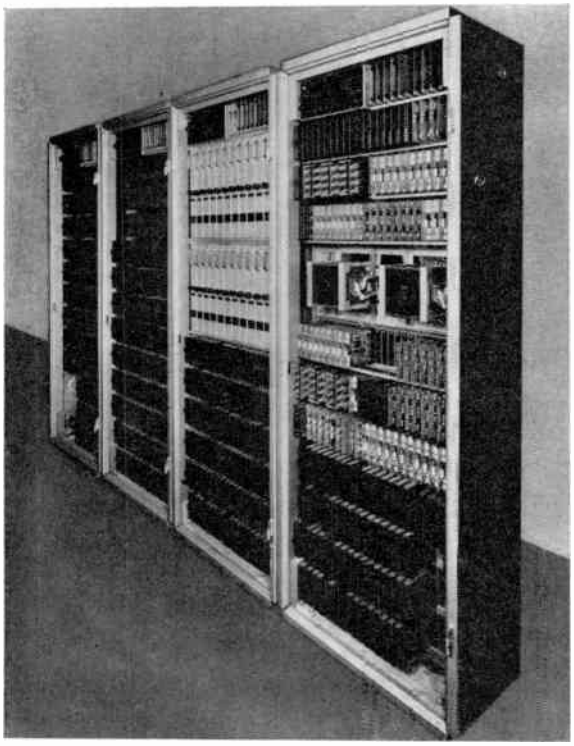


Fig. 6 (*b*). Computer of similar capacity to Fig. 6 (*a*), but using conventional printed circuits.

capabilities of future high-speed logic units, the dispersal of assemblies within the computer must be the minimum consistent with an engineering form which provides for ease of servicing. Other things being equal, the compression of a much greater number of circuit elements into a given volume would result in increases in temperature. But these are more than offset by the use of silicon devices in place of germanium, and therefore air-cooling with normal low-pressure fans will still be adequate. A physically small high-speed computer has been constructed using the silicon micro-logic devices with a stage-delay time of 5 nanoseconds described earlier. It includes a store of 4096 24-bit words with a cyclic time of 1·2 microseconds. In order to achieve maximum reliability diode-transistor logic with its greater noise immunity was used, cheapness being achieved by using the simplest forms of logical design and organization.

Coming next to the memory, the possibilities of storing an increasing amount of information in a decreasing volume have already been discussed. With earlier computers access to the stored information was only on a cyclic basis. Therefore, a machine capable of working quickly had to have a very short cyclic time, and this meant that the designer would always be battling with the restrictions of a Maxwell-limited approach. Nature side-tracks speed limits

imposed by Maxwell's equations. As in electronic computers, she conveys information with binary digits, but does so at a transmission speed of no more than 60 miles/hour. The very rapid transfer of large amounts of information is obtained by using multiple-access on a tremendous scale to the store, connection of this store to the various types of peripheral equipment—the sensors—being by multicore cables, the thousands of fibres which make up the nerves. (Consider how quickly one can take in and recollect a scene.)

Such considerations point to the need for very large electronic stores with facilities for multiple random access and, indeed, such devices are the objectives of much current work. A particular problem that springs to mind is the Post Office Savings Bank, where the application of electronic data processing to the accounting system would require automatic access to any one of 22 million accounts.

2. *System Organization.* Both the engineering form and the system organization of the next family of computers should be much more flexible than have been the designs of the past. Apart from machines which are installed to meet a specific and unvarying operational requirement—and this may hold for some industrial applications—we should be able to offer the user a system capable of being adapted and enlarged to meet changing needs.

All data-processing systems are likely to become more fully automatic. For example, the loading and unloading of magnetic tape will become less necessary with very large random-access stores where all the data held in the system will be accessible at all times. Routine operations, for which a human operator is now necessary, will be progressively reduced in number. Programs may be drawn up automatically and work schedules evaluated by the machine itself. Peripheral devices will be switched on and off automatically, and the computer may be arranged to make automatic connection to data transmission links and signal its requirements for more information as it approaches the end of a task.

An important trend is already apparent towards making computers more 'self-conscious'. Built-in facilities enabling computers to detect overflow in their arithmetic units, faulty or dubious blocks of input or output data and store-parity errors are becoming commonplace. In so far as engineering operation is concerned, automatic routining facilities—dear to the telephone engineer—may well be included in the electronic circuitry, enabling the computer to identify circuits, or circuit elements, which are not operating within prescribed limits. Automatic transfer of the working load to other circuits is a more distant possibility.

The Computer in Relation to Control Processes

In many industrial applications the peripheral equipment on the output side may control the situation without intervention by an operator. The guidance of a space vehicle in flight is an example, information being fed into the system from a tracking radar. In such cases the whole system becomes a closed servo loop, and this is true of many industrial applications. In other situations, and generally when the computer is used for business or commercial applications, the loop is broken and someone has to act on the data presented to him. Equally, he, or someone else, has to determine what data are made available to the computer.

But as the trend progresses towards making data-handling and control systems more automatic and less a question of conscious concern to an operator, the remaining decisions which the latter will have to take will be harder to take correctly as they become fewer. The effect of a bad decision will be more calamitous. This seems to be a feature of all complex and highly automated systems, and one can think of no reason why computers should be an exception.

The data may arrive from a variety of sources. With business and commercial applications it was convenient at first to use punched cards or tape, because they were existing and suitable means of mechanically recording data, but these are now being replaced by magnetic tape. With industrial applications 'on-line' working of the computer may be practicable, but this at some point will involve the translation into coded impulses of physical measurements, e.g. of temperature and flux readings from the interior of a nuclear reactor, signals derived from search radar in an air-traffic control system, readings of strain gauges distributed over a structure, etc. Constant checking is needed at all stages to ensure that sensory data are not being lost or distorted in transmission, and that control signals are getting through and being acted upon. Moreover, there must be a safeguard to ensure that no fatally extreme instruction results in unchecked action, even though the instruction has been worked out quite correctly by the control program. In some situations, e.g. blind landing of aircraft, this might be very important.

In this connection it is interesting to note that work has been done on the development of what has been described as a 'learning matrix'. This is a special form of associative store in which information may be accumulated as the average of a number of patterns applied successively. The engineering feasibility of the 'learning matrix' has been demonstrated and its applications in control engineering are obvious.

In many of these applications the computer is at a distance from the point of origin of the data, and

Fig. 7. Alpha-numeric display of processed information on cathode-ray tube.

these have to be transmitted over cables or radio links. The amount of data that can be transmitted in a given time depends not only on the frequency bandwidth of the communication channel, but also on the complexity of the built-in checking system. Transmission errors are usually detected by redundant coding and correction is effected by re-transmission of the corrupted part of the information. As an example of what is possible, a system designed to meet the internationally agreed transmission speeds will transmit data over normal telephone lines at a rate of 100 characters every second with an anticipated accuracy of not more than one error in 10- or 100-million characters. The greatest efficiency will always be obtained by planning data-processing and the data transmission as an integrated whole.

Input and Output Arrangements

Communication between the computer and the operator is an outstanding problem. This concerns the input and output arrangements, it being reasonably certain that the capabilities of the electronic computer itself will continue to increase comfortably in advance of those of the peripheral equipment with which it is associated.

The majority of printing, forming the output of present-day computing systems in business offices, comes from electro-mechanical, line-at-a-time printers. Printers with a capacity of 80, 120 or 160 columns at speeds of 1000 lines per minute are available.

In many industrial applications printers are obviously not appropriate. For such cases a computer output device has been developed which is fully compatible with computer operating speeds and with the rate human operators can absorb information. In this device lines of data are written in sequence on the screen of a cathode-ray tube. Each character is written in turn by deflecting the electron beam to the position where the character is required and then applying a fast low-amplitude deflection to trace out the letter or figure. Each character is written in a period of 20 microseconds, so that the system is capable of writing 50 000 characters per second. It is necessary to repeat the information at a recurrent frequency high enough to avoid flicker effects, but the cathode-ray tube used has an after-glow characteristic such that it retains a steady picture if repeated at more than 10 times per second. The store capacity in the associated computer is arranged to match this rate, and continuous cycling occurs at approximately 12 times per second. The storage capacity and amount of information displayed considerably exceeds that which can be used quickly by one man and makes it possible to feed a number of individual operating positions, if desired, with completely different messages.

Two Examples of Application

The possible applications of computers are legion. I intend to mention only two. The first is a possible accounting application, which lies close to what were my own principal interests. All long-distance telephone calls handled by operators are brought to account by a ticket which the operator completes and which forms the first step in a chain of processes, ending in a bill to the subscriber. This is an expensive process and, with the introduction of routing and switching equipment to enable subscribers to dial their own long-distance calls, there is anyhow no regular need for an operator. The British Post Office has therefore done away with accounting tickets for subscriber-dialled calls which are recorded by forward-stepping of a meter in the exchange individual to the calling subscriber; the number of units recorded is proportional to the distance and duration of the call. At quarterly periods all the meters—there may be several thousand—in an exchange are read to give the number of units to be charged to the subscriber. This use of personnel appears somewhat out of pattern with fully-automatic telephone switching systems, particularly against a background of anticipated progress from the present electromechanical switching equipment to equipment that is fully electronic. It should be possible to devise a system in which the pulses that are used at present to step the electromechanical meters will be used to up-date the store in an electronic data-processing system. At appropriate intervals this would produce complete accounts to all the subscribers.

The second application is an industrial one and relates to the introduction of computers into large electricity generating stations. I shall draw on work undertaken for a 1180 MW station being built for the Central Electricity Generating Board and in which the power source is a gas-cooled, graphite-moderated reactor system. For such stations operation and control techniques, which are merely extensions of those used for many years, are outdated. In a nuclear station there is very extensive additional instrumentation of the reactor and the amount of information that has to be taken into account is much greater than that in the operation of a conventional power station. Therefore, it becomes desirable to re-think the problem and plan a system that will enable the control engineer to get immediate information on the state of a section of the station by simply touching a button on the central control desk. This information will be displayed in tabular form on the screen of a cathode-ray tube. In this particular case it will be gathered from a total of 3940 prime measurement and state devices in the reactor, associated turbo-alternator sets and other parts of the station. Before display the information will be processed by a high-speed digital computer which is capable of analysing it to determine whether the plant is operating satisfactorily, detecting any potentially dangerous trends before they become serious, and informing the operator as necessary. It would be further possible for the computer itself to make decisions and take direct control action if required.

Conclusions

I do not intend to discuss other applications. Obviously, they will touch every aspect of our daily lives. Even translation from one language to another is within the list of possible uses for computers. Government data-processing will increase as demands for more social concern for each individual require more to be known about him during his entire life. Computers will be essential to control the traffic of the future—in the air and in the cities. Computer programmed machines in factories will work untended. A long time ago Aristotle wrote: "When looms weave by themselves men's slavery will end". But this is a soulless electronic monster we have created, his logic circuits knowing no language but 'yea' or 'nay', and he will create some social problems as he comes among us. Sir Leon Bagrit may deal with them in his forthcoming B.B.C. Reith Lectures.

I cannot conclude without expressing my very sincere thanks to all my colleagues who have provided me with ideas and information during my preparation of this lecture. It has been a stimulating experience. I must apologise to some of them because I have not followed up all the suggestions they made. If I had we should have been here much longer.

There is one other thing I should like to say. This is an extremely fast-moving technology, particularly as it enters micro-electronics. In the public mind it is often associated with American development. But all I have described tonight has been based on British achievement—by one organization or another—and I am convinced that if we utilize our ability aright we can hold our own in this as in other fields.

(Address No. 34)

U.D.C. 621.382

Analogue Multiplication with the Space-charge-limited Surface-channel Triode

By

G. T. WRIGHT, D.Sc.†

Summary: It is shown that the space-charge-limited surface-channel triode can be operated so as to give a current-voltage characteristic described by $i = 4PV_1V_2$, where P is perveance, V_1 is gate voltage and V_2 is drain voltage. The accurate analogue multiplication of two voltages with a cadmium sulphide thin-film triode is described. Materials and design problems are briefly discussed and it is concluded that by using a silicon triode it should be possible to achieve accurate multiplication at frequencies up to several Mc/s with a dynamic range of about two orders of magnitude for each input signal.

1. Introduction

The operating mechanisms of surface-channel solid state triodes, as originally proposed by Heil,[1] have recently been described. In particular, Ihantola[2] has considered the situation when the conducting channel consists of semiconductor material, and Wright[3, 4] has considered the situation when the conducting channel consists of high resistivity material. In the latter case current is carried by space-charge and operation occurs under space-charge-limited conditions. The analysis of device characteristics[4] shows that the space-charge-limited current between source and drain is given by

$$i = PV_2(2V_1 - V_2) \qquad \text{.....(1)}$$

where V_1 and V_2 are gate and drain voltages respectively ($V_1 \geqslant V_2$) measured relative to the source at zero (earth) potential, and the perveance P is equal to $\epsilon\mu w/2hd$, where w is the length of the electrodes, h is the thickness of the gate insulation layer and d is the spacing between source and drain ($w \gg d \gg h$). The permittivity of the gate insulation is ϵ and the mobility of charge carriers in the conducting channel is μ.

The current-voltage characteristic described by eqn. (1) has been observed in experimental devices by Wiemer *et al.*,[5] by Hofstein and Heiman[6] and particularly by Zuleeg.[7]

2. Multiplication

A more general description of device characteristics can be obtained from (1) by allowing for application of a voltage V_0 to the source ($V_0 \leqslant V_1$). In this case, with voltages V_0, V_1 and V_2 applied to source, gate and drain respectively, we have

$$i = P(V_2 - V_0)[2(V_1 - V_0) - (V_2 - V_0)] \qquad \text{.....(2)}$$

Various interesting functional relations between current and the applied voltages can be derived from this equation. For instance, if $V_1 = V_2$ and $V_0 = 0$ we obtain

$$i = PV_1^2 \qquad \text{.....(3)}$$

This is, of course, the straightforward diode connection with source earthed and gate and drain linked together. If $V_1 = V_2$ and $V_0 \neq 0$ we obtain

$$i = P(V_1 - V_0)^2 \qquad \text{.....(4)}$$

representing the square of a sum or difference. In particular, if $V_0 = -V_2$ and $V_1 \neq 0$ we obtain

$$i = 4PV_1V_2 \qquad \text{.....(5)}$$

representing multiplication.

The current-voltage characteristic described by eqn. (5) is of considerable interest because multiplication is an analogue process which is difficult to carry out with speed and accuracy. It is worth while therefore to consider the use of the space-charge-limited surface-channel triode for this purpose.

A practical demonstration of multiplication using a cadmium sulphide thin-film triode is shown in

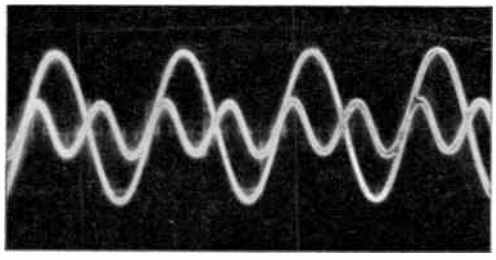


Fig. 1. Showing the amplified multiplier a.c. output signal component superimposed on the a.c. input signal component. The conditions of operation are such that the first-harmonic product terms cancel; the output thus consists only of the small second-harmonic term plus all the error terms contained in the large first-harmonic terms. The lack of distortion, even in the small second-harmonic term, shows that accurate multiplication has been achieved.

† Electronic and Electrical Engineering Department, University of Birmingham.

Fig. 1. This demonstrates the multiplication of two voltages $V_1 = A + a \sin \omega t$ and $V_2 = B - b \sin \omega t$ according to the equation:

$$(A + a \sin \omega t)(B - b \sin \omega t)$$
$$= AB - \frac{ab}{2} + (aB - bA) \sin \omega t + \frac{ab}{2} \cos 2\omega t$$

The steady bias voltages A and B were of order 1 V. The alternating voltages $a \sin \omega t$ and $b \sin \omega t$ were derived from a 1 kc/s oscillator and were of amplitude about 0·1 V. In order to demonstrate the accuracy of multiplication the bias voltages A and B were adjusted so that $aB = bA$. If accurate multiplication can be achieved the large first-harmonic products should balance and then neutralize each other, leaving only the small second-harmonic term. The diagram shows the amplified multiplier output signal under these conditions superimposed on the a.c. input signal. It can be seen that there is very little distortion in the output, even under these stringent test conditions.

3. Discussion

The experimental demonstration shown in Fig. 1 indicates that the space-charge-limited surface-channel triode can be used for accurate analogue multiplication. Used in the manner described it is limited, however, by the operating requirement that the gate voltage should always equal or exceed the drain voltage. If this condition is violated current is less than given by eqn. (5) and accuracy is lost. It is probable, of course, that many applications exist which do not violate this condition. It can be removed, however, by arranging for the two signal voltages to be compared at the input and switched so that the larger is always applied to the gate and the smaller to the drain. Alternatively two triodes can be used with the signal input connections interchanged so that at all times one of the triodes is operating in the correct mode. Selection of the appropriate triode can always be made on the basis that its current is the larger. In this connection it is of relevance to note that the functional relation between current and voltage given by eqn. (1) is not affected by the triode dimensions. Thus accurate matching of triode characteristics can be achieved by the use of external current or potential dividers or amplifiers.

A suitable material for device fabrication is high-purity silicon. The technology of this material is well understood and it is known to produce reliable devices. Silicon is available in single crystal form with resistivity greater than 10^4 ohm-cm. This purity is sufficient to ensure that thermal generation of carriers is small enough so that shunt conduction between source and drain is negligible. Operation then takes place under space-charge-limited conditions as is required for the validity of eqn. (1).

The design of surface-channel triodes for use as multipliers involves the usual conflicts between dynamic range, speed and accuracy. Detailed discussion of the various relevant factors requires analysis of a specific design and is beyond the scope of this discussion. However, it is worth while to indicate some of the important considerations. At small drain voltages of the order of a few times kT/e, carrier diffusion is significant as a current mechanism and current is greater than indicated by (1). At large drain voltages field dependence of mobility causes the perveance P to decrease and current will be less than indicated by (1). Limitations on the gate voltage are less severe, being restricted at large voltages by the possibility of breakdown of the gate insulation layer. Between the permissible extremes of small and large drain voltages there should exist a dynamic range of at least two orders of magnitude of input signal through which eqn. (5) should be valid.

A further limitation concerns the permissible level of power dissipation in the device. The instantaneous level of power dissipation is

$$W = 8PV_1V_2^2 \qquad \text{.....(6)}$$

and obviously rises very rapidly indeed if both signal voltages increase together. In the majority of practical applications, however, the mean power level is much less than the peak and providing that the source can supply current surges adequately the mean power level is probably the better limit to adopt. With an electrode length $w = 1$ cm, a spacing between source and drain of $d = 0{\cdot}02$ cm, a thickness for the gate insulation layer of $h = 1$ micrometre and using typical materials parameters of $\epsilon = 5 \times 10^{-11}$ F/m and $\mu = 0{\cdot}1$ m^2/volt-second we obtain

$$i = 0{\cdot}5\, V_1V_2 \text{ mA} \qquad \text{.....(7)}$$

If a mean power dissipation of 1 watt is regarded as tolerable then a working input voltage range from about 0·1 V to about 10 V is indicated.

The speed of response is controlled essentially by the transit time of carriers between source and drain. This transit time is given by

$$t = \frac{4d^2}{3\mu(2V_2)} \qquad \text{.....(8)}$$

With the device dimensions and materials parameters given above we obtain $t = 27$ mμs when $V_2 = 10$ V. This suggests that when working with large signal levels accurate multiplication can be achieved at frequencies up to several megacycles. Operation at higher speeds can be obtained, of course, by reducing the spacing between source and drain. To maintain accuracy it is necessary also to reduce the thickness of the gate insulation layer to maintain the condition $d \gg h$. Current therefore increases considerably, and in order to avoid excessive power

dissipation the dynamic range of the device is reduced. If operation at high speeds is not essential the use of low mobility carriers enables operation to take place with smaller current output and greater dynamic range.

In addition to the considerations outlined above, imperfections in materials, such as the existence of interfacial states along the conducting channel, and in design, such as the existence of fringe fields at the electrode edges, may produce deviations from the ideal characteristic of eqn. (5). Further research is necessary fully to evaluate and control mechanisms of operation and to achieve optimum design. Nevertheless, the surface-channel space-charge-limited triode would seem to offer good prospects for fast and accurate analogue multiplication.

4. Acknowledgments

It is a pleasure to acknowledge that the surface-channel triode used was supplied by R. Zuleeg and the demonstration of multiplication was devised and constructed by P. W. Webb.

5. References

1. O. Heil, Brit. Patent, No. 439457, 1935.
2. H. K. J. Ihantola, "Design theory of a surface field-effect transistor", ASD Technical Note 61–133, 1961.
3. G. T. Wright, "Space-charge-limited solid state devices" *Proc. Inst. Elect. Electronic Engrs*, **51**, p. 1642, 1963.
4. G. T. Wright, "Theory of the space-charge-limited surface-channel dielectric triode", *Solid State Electronics*, **7**, p. 167, 1964.
5. P. K. Wiemer, F. V. Shallcross and H. Borkan, "Coplanar-electrode insulated-gate thin-film transistors", *R.C.A. Review*, **24**, p. 661, 1963.
6. S. R. Hofstein and F. P. Heiman, "The silicon insulated-gate field-effect transistor", *Proc. I.E.E.E.*, **51**, p. 1190, 1963.
7. R. Zuleeg, "Electrical evaluation of thin-film cadmium sulphide diodes and transistors", *Solid State Electronics*, **6**, p. 645, 1963.

Manuscript first received by the Institution on 18th March 1964 and in final form on 8th June 1964. (Contribution No. 79.)

INSTITUTION NOTICES

Birthday Honours List

The Council congratulates the following member of the Institution whose appointment appears in Her Majesty's Birthday Honours List:

William Kenneth Newson (Member) to be a Member of the Most Excellent Order of the British Empire. (Mr. Newson, who has been with the British Broadcasting Corporation for 40 years, has held the position of Engineering Recruitment Officer since 1953. He is also a member of the Institution's Education and Training Committee.)

Other Birthday Honours were published in the June issue of *The Radio and Electronic Engineer*.

Ghana Public Service Commission

It has recently been announced that the Ghana Public Service Commission now accepts Graduate and Corporate Membership of the Institution for professional engineering appointments in the Ghanaian Civil Service. This decision, which will be welcomed by members in Ghana, affects appointments in the Posts and Telegraphs Department and the Ghana Broadcasting Corporation.

Rhodesian Institution of Engineers

All Graduate and Corporate Members of Institutions on the Engineering Institutions Joint Council are invited to apply for membership of the appropriate grade in the Rhodesian Institution of Engineers. The standards of entry for the various grades are parallel to those required by professional Institutions in the United Kingdom and their examinations are accepted by the Rhodesian Institution.

The Rhodesian Institution of Engineers was established by Act of the Registered Assembly of Southern Rhodesia and its aims are similar to those of professional engineering Institutions in the United Kingdom, and other Commonwealth countries.

Members of the I.E.R.E. who are resident in Southern Rhodesia and wish to join the Rhodesian Institution of Engineers should write to P.O. Box 660, Salisbury, Southern Rhodesia.

Institution Group in Nigeria

The Council has received a request from Corporate Members in Nigeria for permission to establish a Group for the purpose of holding technical meetings. Several members have formed an informal committee and are already planning programmes for the coming months. Offers of support for these activities will be welcomed and further particulars may be obtained from Mr. P. E. Olisa, A.M.I.E.R.E., Principal, Posts and Telegraphs Training Centre, Lagos.

Symposium on Microminiaturization

A limited number of preprints of the Joint I.E.E.–I.E.R.E. Symposium on "Microminiaturization", held in Edinburgh in April 1964, are still available at 30s. per set. Inquiries should be addressed to Mr. R. D. Pittilo, 35 Crawfurd Road, Burnside, Rutherglen, Glasgow.

Conference on Electron Emission

The Institute of Physics and the Physical Society will be holding a Conference on Electron Emission at the University of Keele, Staffordshire, on the 1st and 2nd October 1964. It is proposed to include the following review papers: Photo-cathodes; Secondary Emission; Alternatives to Thermionic Emission; Recent Research on Secondary Electron Emission at Mainz; Electron Emission by Fast Ions.

Advance registration will be necessary; the closing date will be 28th August. Further details and application forms may be obtained from the Administration Assistant, The Institute of Physics and the Physical Society, 47 Belgrave Square, London, S.W.1.

Meetings on Quality and Reliability

The National Council for Quality and Reliability is organizing a two-weeks' residential course on "Quality and Reliability Principles and Practice" at the College of Aeronautics, Cranfield, 6th–18th September 1964.

The N.C.Q.R. is also arranging, in conjunction with the Institution of Mechanical Engineers, a Colloquium on Reliability to be held in London on 10th and 11th November 1964.

The I.E.R.E. is a member of N.C.Q.R. and further details and application forms may be obtained from 8–9 Bedford Square, London, W.C.1.

June *Journal* 1964

Members and subscribers who bind their own *Journals* are asked to note that a mistake occurred in the collation of the last section of the June issue. This prevents the text section (pp. 465–472) from being effectively separated from the advertisement pages. The text pages have been reprinted and a copy may be obtained on application to the Publications Department, I.E.R.E., 9 Bedford Square, London, W.C.1.

Index to Volume 27

The June 1964 issue completed Volume 27 of *The Radio and Electronic Engineer* which covers the period January–June 1964. An Index to the volume will be circulated with the August issue.

U.D.C. 621.372.512.3: 621.319.5: 539.1.07

An Analysis of Stagger-tuned Pulse Receiver Systems with special reference to N.M.R. Relaxation-time Measurements

By

H. PURSEY, B.Sc.†

Summary: A 30-Mc/s receiver is described for use in pulsed nuclear magnetic resonance apparatus. The receiver is designed to have a good signal/noise ratio, as well as rapid recovery from a pulse overload. Various types of stagger-tuned arrangements are examined theoretically, and it is shown that for optimum results the poles of the amplifier transfer function should coincide with the roots of a Bessel polynomial. An amplifier designed on this principle is described, and is shown to have characteristics which are in agreement with theoretical predictions.

1. Introduction

Several authors have published descriptions of equipment for measuring the relaxation times of nuclear magnetic resonance processes by the spin-echo method.[1,2,3] Briefly, the method consists of applying a short, high-power pulse of radio-frequency energy to the sample under test, and observing the decay of the signal so excited. To prevent the exciting pulse passing directly to the receiver either a crossed-coil system or a balanced bridge may be used, but in either case a proportion of the transmitted energy will reach the receiver because of imperfect balancing; this is particularly the case when very short pulses are used, partly because of the high power needed to turn the nuclear magnetic moments through a given angle in a short time, and also because of the wide frequency distribution of energy, it being difficult to balance the bridge or crossed-coil system effectively over a wide band of frequencies.

The problems of designing a receiver which will recover rapidly from a large pulse overload have been encountered in receivers for radar systems.[4] In this connection the critical parts of the receiver are the interstage coupling circuits, the cathode bias circuits, and the high-tension decoupling circuits. For a 30-Mc/s receiver the time-constants of the cathode-bias circuits need not exceed 0·2 μs and the high-tension supply can be derived from a low-impedance stabilized source, individual stages being decoupled by ferrite beads. This leaves the question of interstage coupling to be considered in more detail.

2. The Interstage Coupling Circuits

To avoid the time-constants associated with resistance-capacitance networks it was decided that transformer coupling would be used, and the question then arises as to whether one or both of the windings should be tuned. This in fact will depend on the number of stages in the receiver, which may reasonably be fixed at six to achieve an overall gain of the order of 120 dB with a gain of 20 dB per stage. One tuned circuit per stage should then give adequate selectivity, whereas introducing more tuned circuits would slow down the transient response without any compensating advantage. The receiver to be described employs untuned-primary/tuned-secondary transformers throughout, the stage gain being given by

$$\eta = -kg_mR_L/n$$

where η is the gain per stage (voltage ratio),

k is the primary-to-secondary coupling coefficient,

g_m is the slope of the preceding tube,

R_L is the secondary load resistance,

n is the ratio of secondary to primary turns.

To ensure that the primary resonant frequency is well outside the receiver pass-band, while maintaining a reasonably high stage gain, n is made equal to 2, and the secondary inductance is made as large as possible so that R_L is large for a given Q-factor. Here a compromise is necessary, since if the coil inductance is made too large it will resonate with only the circuit stray capacitances. Such an arrangement would have poor long-term stability, and in practice the coils are designed to resonate with capacitors of about 30 pF. The circuits are loaded with shunt resistors of several thousand ohms, which are adjusted experimentally to give the desired Q-values in conjunction with the additional loading imposed by the input conductance of the following stage.

3. Stagger Tuning

Although a satisfactory receiver can be built using six identical tuned circuits, it is possible by means of a suitable stagger-tuning arrangement to obtain an improved transient response for the same overall gain. In principle, a stagger-tuned amplifier has the poles of its tuned circuits arranged in a regular pattern to give a desired frequency response, the poles being defined as the points in the complex frequency plane

† National Physical Laboratory, Basic Physics Division, Teddington, Middlesex.

at which the response of the corresponding circuit is infinite. Thus, if the resonant frequency and Q of the ith circuit are respectively $\omega_i/2\pi$ and Q_i, then the pole location in the s-plane is given by

$$s = s_i = -(\omega_i/2Q_i)+j\omega_i \qquad \text{.....(1)}$$

The pole arrangement is generally based on a low-pass filter response, those commonly used being the Butterworth maximally flat response, the Chebyshev equal-ripple response, and the Bessel maximally flat phase response.[5,6] Our purpose is to investigate the transient behaviour of these filters and so decide which is the most suitable for a spin-echo amplifying system.

The Butterworth response is derived from the function

$$f(x) = (1+x^{2n})^{-\frac{1}{2}} \qquad \text{.....(2)}$$

The poles of this function are the $2n$ values of $(-1)^{1/2n}$, which are equally spaced around the circle of unit radius with its centre at the origin. x is taken to represent frequency, so that if x_i is the ith pole the corresponding pole in the s-plane is given by $s = s_i = 2\pi j x_i$. Poles in the lower half of the x-plane correspond to poles in the right-hand half of the s-plane, indicating an oscillatory system. We therefore include only those poles which lie above $\mathrm{Im}(x) = 0$; symmetry shows that for real values of x this is equivalent to taking the square root of the response arising from the complete pole set, which leads us to the response given by eqn. (2) above.

As n tends to infinity, $f(x)$ approaches the response of an ideal low-pass filter, having the value unity for $|x| < 1$ and zero for $|x| > 1$. It may also be shown to represent the response of a physically realizable network, the poles being confined to the upper half-plane in x, and located either on the positive imaginary axis, or in symmetrical pairs on either side of it.[6]

To obtain a band-pass characteristic the cluster of poles is moved along the real axis of x and centred at $x = x_0$, where $x_0 > 1$. In order to conform to the conditions of realizability a similar set of poles must be centred on $x = -x_0$. Such a system will have a response which is symmetric about $x = 0$, but not about $x = x_0$, although the latter asymmetry will be slight provided $x_0 \gg 1$, as is often the case. A further modification will be to locate the poles on a semicircle of radius other than unity, corresponding to a variation in the pass-band of the amplifier.

A Chebyshev response is obtained by modifying the circle locating the Butterworth poles to an ellipse whose major axis lies along the real axis in the x-plane, the corresponding response function being

$$f(x) = \{1+\varepsilon^2 C_n(x)\}^{-\frac{1}{2}} \qquad \text{.....(3)}$$

where ε is a real constant lying between 0 and 1, and $C_n(x)$ is defined by the equation

$$C_n(x) = \cos(n \arccos x) \qquad \text{.....(4)}$$

The term 'equal-ripple response' derives from the fact that in the pass-band, where $|x| < 1$, $f(x)$ oscillates between $(1+\varepsilon^2)^{-\frac{1}{2}}$ and $(1-\varepsilon^2)^{-\frac{1}{2}}$.

The Butterworth and Chebyshev poles for a low-pass filter are given by the following formulae:

$$s_k = \sigma_k + j\omega_k \qquad \text{.....(5)}$$

where
$$\sigma_k = \sin(2k-1)\pi/2n \qquad \text{.....(6)}$$
(Butterworth)

$$\sigma_k = \tanh a \sin(2k-1)\pi/2n \qquad \text{.....(7)}$$
(Chebyshev)
$$(a = n^{-1} \operatorname{arc\,sinh} \varepsilon^{-1})$$

$$\omega_k = \cos(2k-1)\pi/2n \qquad \text{.....(8)}$$
(Butterworth and Chebyshev)

A suitable amplifier for spin-echo measurements will have a bandwidth of ± 1 Mc/s centred on 30 Mc/s. Then if $s_0 = j\omega_0 = 60\pi j$, we have for a Butterworth amplifier:

$$s_k = s_0 - 2\pi\{\sin(2k-1)\pi/12 + j\cos(2k-1)\pi/12\} \qquad \text{.....(9)}$$

and for a Chebyshev amplifier:

$$s_k = s_0 - 2\pi\{\tanh a \sin(2k-1)\pi/12 + j\cos(2k-1)\pi/12\} \qquad \text{.....(10)}$$

The Chebyshev amplifier given by eqn. (10) will have a higher mid-band gain than the corresponding Butterworth amplifier, since the Chebyshev poles are closer to s_0 than the corresponding Butterworth poles. To make a meaningful comparison between the two amplifiers we normalize the Chebyshev poles so that the amplifiers have equal mid-band gains. This corresponds to an increase in the pass-band of the Chebyshev amplifier. The Chebyshev poles are then given by

$$s_k = s_0 - 2\pi c\{\tanh a \sin(2k-1)\pi/12 + \cos(2k-1)\pi/12\} \qquad \text{.....(11)}$$

where c is chosen so that

$$\prod_{k=1}^{6}(s_k - s_0) = (2\pi)^6 \qquad \text{.....(12)}$$

as is evidently the case for the Butterworth amplifier.

The pole locations for the Bessel maximally flat phase response amplifier are the solutions of the equation $B_n(s) = 0$, where $B_n(s)$ is the nth order Bessel polynomial, which may be defined by the recursion formula

$$B_n = (2n-1)B_{n-1} + s^2 B_{n-2} \qquad \text{.....(13)}$$
$$B_0 = 1$$
$$B_1 = 1+s$$

In particular

$$B_6(s) = 10395 + 10395s + 4725s^2 + 1260s^3 + 210s^4 + 21s^5 + s^6 \qquad \text{.....(14)}$$

The pole locations for the three amplifiers, normalized for equal gain at the centre frequency of 30 Mc/s, are given in Table 1.

To obtain the practical circuit parameters we use eqn. (1) above:

$$s_k = -\omega_k/2Q_k + j\omega_k$$

where $\omega_k = 2\pi f_k$.

Thus
$$f_k = (2\pi)^{-1}\,\mathrm{Im}(s_k) \qquad \text{.....(15)}$$

$$Q_k = -\mathrm{Im}(s_k)/2\,\mathrm{Re}(s_k) \qquad \text{.....(16)}$$

For the Bessel amplifier whose pole locations are given in Table 1 we have:

	1	2	3	4	5	6
f(Mc/s)	29·04	29·44	29·81	30·19	30·56	30·96
Q	27·85	18·76	16·49	16·49	18·76	27·85

The transfer function of the amplifier is given by the formula

$$\Phi(s) = As^6 / \prod_{i=1}^{6} (s-s_i)(s-\tilde{s}_i) \qquad \text{.....(17)}$$

where A is a constant which may conveniently be chosen so that the gain of the amplifier is unity at the centre of the pass-band, and $\tilde{s}$ is the complex conjugate.

The frequency response of the amplifier may be obtained by considering the steady-state component of the response to a function $f(t) = \exp j\omega t$ applied at $t = 0$. The Laplace transform of $f(t)$ is

$$\mathscr{L}f(t) = \int_0^\infty \exp(j\omega t - st)\,dt \qquad \text{.....(18)}$$

$$= (s-j\omega)^{-1} \qquad \text{.....(19)}$$

and the amplifier output is therefore

$$2\pi F(t) = \int_{c-j\infty}^{c+j\infty} (s-j\omega)^{-1}\,\Phi(s)\exp st\,ds \qquad \text{.....(20)}$$

The steady-state component arises from the pole at $s = j\omega$, so the frequency response is simply

$$\psi(\omega) = |\Phi(j\omega)| = \left|(4\pi\omega)^6 / \prod_{i=1}^{6} (j\omega - s_i)(j\omega - \tilde{s}_i)\right| \qquad \text{.....(21)}$$

where we have put $A = (4\pi)^6$ for normalization.

4. Transient Response of the Amplifier

To calculate the transient response we assume a signal $f(t) = \sin \omega_0 t$ is applied at $t = 0$, where ω_0 is 2π times the mid-pass-band frequency of the amplifier. Since the signal will eventually be rectified and displayed on a cathode-ray tube we are interested in the envelope of the amplifier output; this has been obtained by calculating the output at time corresponding to peaks of the input signal, i.e. at $t = \dfrac{\pi}{2\omega_0} + \dfrac{2n\pi}{\omega_0}$ where n is integral. This procedure is not strictly correct, on account of phase and amplitude distortions which occur during the early part of the transient, so that peaks of the output and input waveforms are not coincident. However, trial calculations show that errors arising from this are negligible when the signal is plotted on a linear amplitude scale.

The Laplace transform of $f(t)$ is given by the formula

$$\mathscr{L}f(t) = \int_0^\infty f(t)\exp(-st)\,dt$$

$$= \omega_0/(s^2+\omega_0^2)$$

$$= \omega_0/(s-s_0)(s-\tilde{s}_0) \qquad \text{.....(22)}$$

where $s_0 = j\omega_0$

Thus, the Laplace transform of the amplifier output is

$$\psi(s) = \omega_0\,\Phi(s)/(s-s_0)(s-\tilde{s}_0) \qquad \text{.....(23)}$$

Inserting the value of $\Phi(s)$ from eqn. (17) we obtain

$$\psi(s) = A\omega_0 s^6 \Big/ \prod_{i=0}^{6} (s-s_i)(s-\tilde{s}_i) \qquad \text{.....(24)}$$

Table 1

	BUTTERWORTH	CHEBYSHEV	BESSEL
s_1	− 1·62621 + j182·426	− 0·751455 + j179·209	− 3·3839 + j182·452
s_2	− 4·44288 + j184·053	− 2·053012 + j181·698	− 5·0245 + j184·963
s_3	− 6·06909 + j186·869	− 2·804467 + j186·006	− 5·7140 + j187·328
s_4	− 6·06909 + j190·122	− 2·804467 + j190·984	− 5·7140 + j189·662
s_5	− 4·44288 + j192·938	− 2·053012 + j195·292	− 5·0245 + j192·027
s_6	− 1·62621 + j194·564	− 0·751455 + j197·781	− 3·3839 + j194·538

and the output as a function of time is therefore

$$F(t) = \frac{A\omega_0}{2\pi j}\int_C \frac{s^6 \exp st\, ds}{\prod_{i=0}^{6}(s-s_i)(s-\tilde{s}_i)} \qquad \text{.....(25)}$$

C being the Bromwich contour from $-j\infty$ to $+j\infty$, indented to the right of singularities along the vertical axis. Poles of the integrand occur at $s = s_i$ and $s = \tilde{s}_i$, giving fourteen first-order poles in the general case when no two stages are alike. In the case of the two staggered triples $s_i = s_{7-i}$ when $i \neq 0$, and we have two first-order poles at $s = s_0$ and $s = \tilde{s}_0$, together with six second-order poles at $s = s_i$ and $s = \tilde{s}_i$, $i = 1$, 2 and 3. Finally, in the case of six identical circuits we have two first-order poles at $s = s_0$ and $s = \tilde{s}_0$, the remaining two poles being of order six. Thus, paradoxically, the simplest electrical arrangement gives rise to the most complex analysis.

To evaluate the integrals arising from eqn. (25) we use Cauchy's formulæ for integration around a pole of order n, since for $t > 0$ the integrand clearly converges in the left-hand half-plane. We obtain the following results:

(1) Six different stages—

$$F(t) = A\omega_0 \sum_{k=0}^{6}\left[\frac{s_k^6 \exp s_k t}{\prod_{\substack{i=0\\ i\neq k}}^{6}(s_k-s_i)\prod_{i=0}^{6}(s_k-\tilde{s}_i)} + \frac{\tilde{s}_k^6 \exp \tilde{s}_k t}{\prod_{i=0}^{6}(\tilde{s}_k-s_i)\prod_{\substack{i=0\\ i\neq k}}^{6}(\tilde{s}_k-\tilde{s}_i)}\right]$$

$$= 2A\omega_0 \operatorname{Re}\sum_{k=0}^{6}\frac{s_k^6 \exp s_k t}{\prod_{\substack{i=0\\ i\neq k}}^{6}(s_k-s_i)\prod_{i=0}^{6}(s_k-\tilde{s}_i)} \qquad \text{.....(26)}$$

(2) Two staggered triples—

$$F(t) = 2A\omega_0 \operatorname{Re}\frac{s_0^6 \exp s_0 t}{(s_0-\tilde{s}_0)\prod_{i=1}^{3}(s_0-s_i)^2(s_0-\tilde{s}_i)^2} +$$

$$+2A\omega_0 \sum_{k=1}^{3}\operatorname{Re}\left[\frac{d}{ds}\frac{s^6 \exp st}{(s-s_0)(s-\tilde{s}_0)\prod_{\substack{i=1\\ i\neq k}}^{3}(s-s_i)^2\prod_{i=1}^{3}(s-\tilde{s}_i)^2}\right]_{s=s_k}$$

$$= 2A\omega_0 \operatorname{Re}\left[-\frac{s_0^6 \exp s_0 t}{(s_0-\tilde{s}_0)\prod_{i=1}^{3}(s_0-s_i)^2(s_0-\tilde{s}_i)^2} + \sum_{k=1}^{3}\frac{s_k^5 \exp s_k t}{(s_k-s_0)(s_k-\tilde{s}_0)\prod_{\substack{i=1\\ i\neq k}}^{3}(s_k-s_i)^2\prod_{i=1}^{3}(s_k-\tilde{s}_i)^2} \times\right.$$

$$\left.\times\left\{6+s_k t - s_k\left(\frac{1}{s_k-s_0}+\frac{1}{s_k-\tilde{s}_0}+2\sum_{\substack{i=1\\ i\neq k}}^{3}\frac{1}{s_k-s_i}+2\sum_{i=1}^{3}\frac{1}{s_k-\tilde{s}_i}\right)\right\}\right] \qquad \text{.....(27)}$$

(3) Six identical stages—

$$F(t) = 2A\omega_0 \operatorname{Re}\left[\frac{s_0^6 \exp s_0 t}{(s_0-\tilde{s}_0)(s_0-s_1)^6(s_0-\tilde{s}_1)^6}\right] + \frac{2A\omega_0}{5!}\operatorname{Re}\left[\frac{d^5}{ds^5}\frac{s^6 \exp st}{(s-s_0)(s-\tilde{s}_0)(s-\tilde{s}_1)^6}\right]_{s=s_1} \qquad \text{.....(28)}$$

To evaluate the second term we proceed as follows:

Let $$\Phi(s) = \frac{s^6 \exp st}{(s-s_0)(s-\tilde{s}_0)(s-\tilde{s}_1)^6}$$

Then

$$\frac{d}{ds}\Phi(s) = \Phi(s)[-(s-s_0)^{-1}+t+6s^{-1}-$$
$$-(s-\tilde{s}_0)^{-1}-6(s-\tilde{s}_1)^{-1}]$$
$$\simeq t\Phi(s)[1-z^{-1}]$$

where $$z = t(s-s_0)$$

provided s, $(s-\tilde{s}_0)$ and $(s-\tilde{s}_1)$ are all large compared with $(s-s_0)$ at s_1, where the derivative is evaluated. We then obtain

$$\frac{d^n}{ds^n}\Phi(s) = t^n\Phi(s)\left[1-\frac{n}{z}+\frac{n(n-1)}{z^2}-\ldots+\right.$$
$$\left.+(-1)^n\frac{n(n-1)\ldots\times 2\times 1}{z^n}\right]$$

so that

$$\frac{d^5}{ds^5}\Phi(s) = t^5\Phi(s)\left[1-\frac{5}{z}+\frac{20}{z^2}-\frac{60}{z^3}+\frac{120}{z^4}-\frac{120}{z^5}\right]$$

and

$$F(t) = 2A\omega_0 \,\mathrm{Re}\left[\frac{s_0^6 \exp s_0 t}{(s_0-\tilde{s}_0)(s_0-s_1)^6(s_0-\tilde{s}_1)^6}\right] + \\ + 2A\omega_0 t^5 \,\mathrm{Re}\left[\Phi(s_1)\left\{\frac{1}{5!} - \frac{1}{4!z_1} + \right.\right. \\ \left.\left. + \frac{1}{3!z_1^2} - \frac{1}{2!z_1^3} + \frac{1}{z_1^4} - \frac{1}{z_1^5}\right\}\right] \quad \text{.....(29)}$$

where $z_1 = t(s_1 - s_0)$

Autocode programs have been written which enable formulæ (21), (26), (27) and (29) to be computed on the N.P.L. ACE computer. Figure 1 shows the impulse responses of six-pole Butterworth, Chebyshev and Bessel amplifiers, compared with an amplifier comprising six identically tuned stages. The superiority of the Bessel arrangement is at once evident. It settles down to within 1% of its final output in 1·3 μs and, unlike the Butterworth and Chebyshev amplifiers, there is no 'overshoot' or 'ringing' after the initial rise.

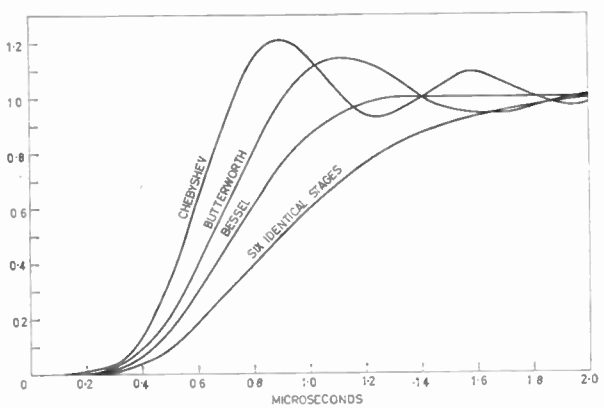


Fig. 1. Comparison of six-pole transient responses.

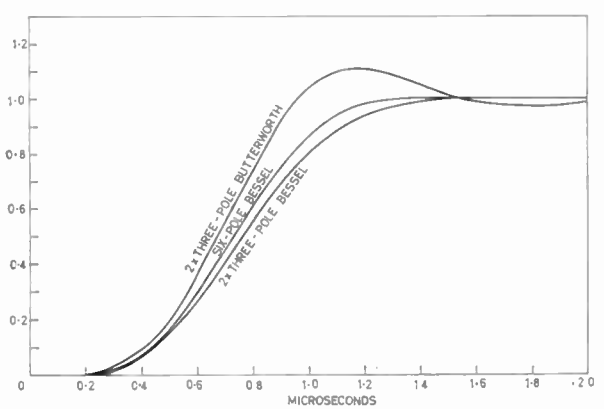


Fig. 2. Comparison of six-pole and 2 × three-pole transient responses.

In Fig. 2 the response of the six-pole Bessel amplifier is compared with an amplifier consisting of two staggered triples, tuned to give firstly a Butterworth and secondly a Bessel response. The six-pole Bessel arrangement is still the best, but is only marginally superior to the Bessel staggered triple. It is interesting to note that the Butterworth staggered triple is better than the six-pole Butterworth amplifier, in that its overshoot is considerably less. The 2 × 3 Butterworth amplifier is commonly used in radar systems,[4] and has in fact been used in the spin-echo system described by Buchta *et al.*[2]

The present amplifier has therefore been designed as a six-pole Bessel filter, using the pole locations given above in Table 1 in the order s_3, s_6, s_2, s_5, s_1 and s_4. The circuit of the amplifier is shown in Fig. 3. The first two stages use high-slope double triodes in cascade configuration, to give a low noise figure for the amplifier, the remaining stages being high-slope pentodes. Gain control is applied via the cathodes of stages 3 and 4, and the amplifier is terminated by a mixer stage, using a 40-Mc/s crystal oscillator, to give an output at 10 Mc/s, the final tuned circuit (at 10 Mc/s) being heavily damped in order not to affect the overall frequency response. An auxiliary output from the 40-Mc/s oscillator is provided so that phase detection can be used if desired; in this case the 30-Mc/s oscillator from which the transmitter pulse is derived would be mixed with the local oscillator in the receiver to give a coherent 10-Mc/s reference signal. To ensure rapid recovery from overload the h.t. decoupling is obtained by using ferrite beads, and cathode bypass capacitors are chosen to give time constants of the order of 0·1 μs.

The superheterodyne arrangement eliminates the possibility of feedback from output and input, which could easily lead to instability in a 'straight' amplifier, and by choosing a relatively low output frequency the problems arising in the design of a second detector are minimized, as will be explained in the following section.

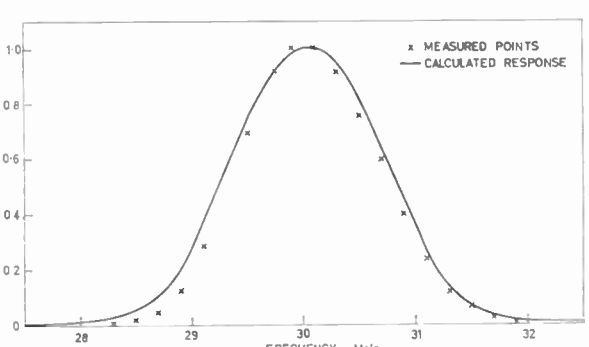


Fig. 4. Frequency response of six-pole Bessel amplifier.

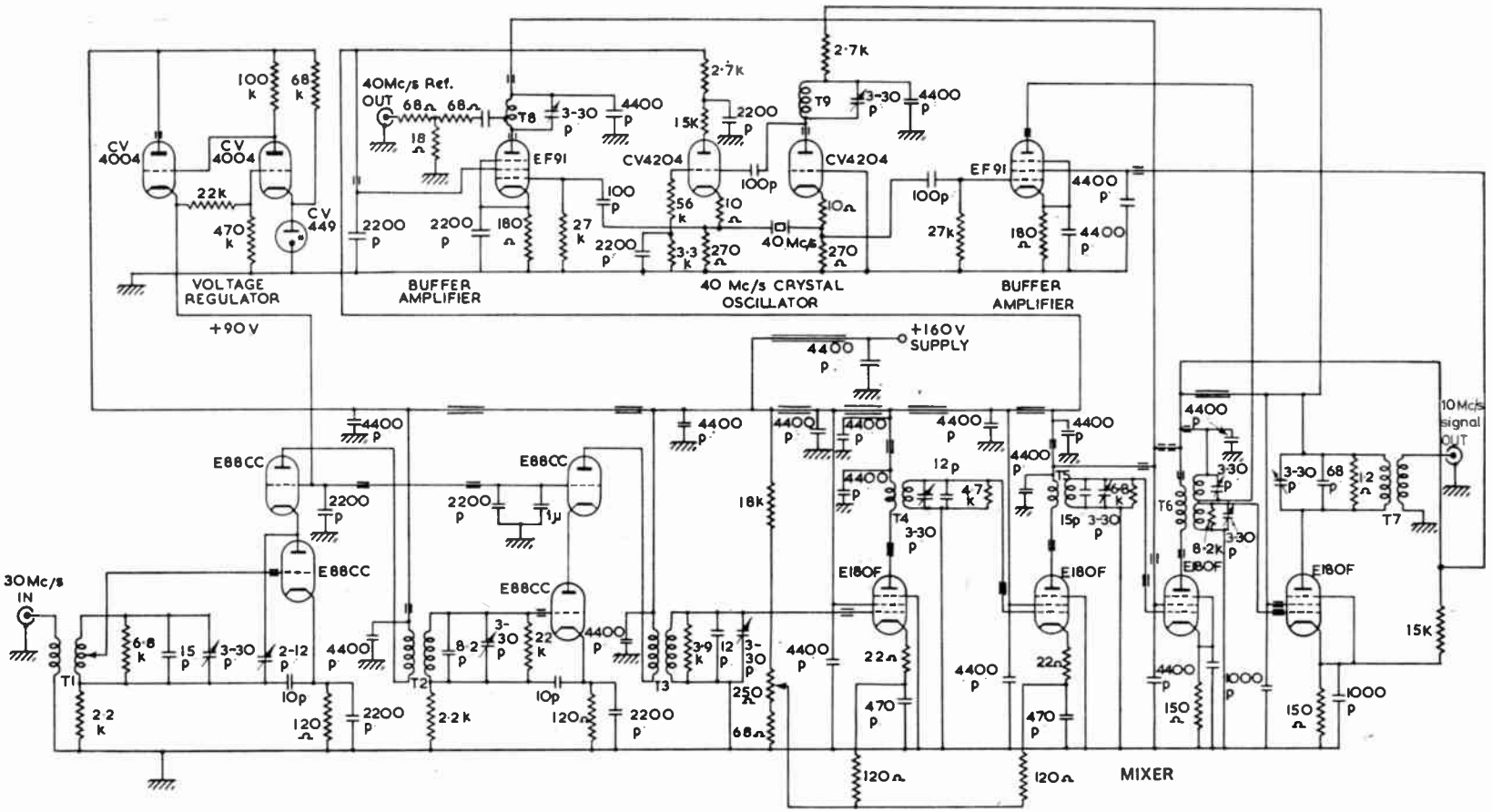


Fig. 3. Circuit of 30-Mc/s amplifier and mixer for spin-echo n.m.r. measurements.

Coil Data

T1 primary 0·1μH secondary 1μH with centre tap. Q 16·5 tuning freq. 29·81 Mc/s.
T2 primary 0·4μH secondary 1μH. Q 28 tuning freq. 30·96 Mc/s.
T3 primary 0·25μH secondary 0·6μH. Q 19 tuning freq. 29·44 Mc/s.
T4 primary 0·25μH secondary 0·6μH. Q 19 tuning freq. 30·56 Mc/s.
T5 primary 0·25μH secondary 0·6μH. Q 28 tuning freq. 29·04 Mc/s.
T6 primary 0·4μH secondary (*a*) 0·475μH (*b*) 0·95μH. Q 16·5 tuning freq. 30·19 Mc/s.
T7 primary 1μH secondary 0·4μH.
T8 0·43μH 4 turns of 32 S.W.G. tap at 1 turn.
T9 0·42μH !

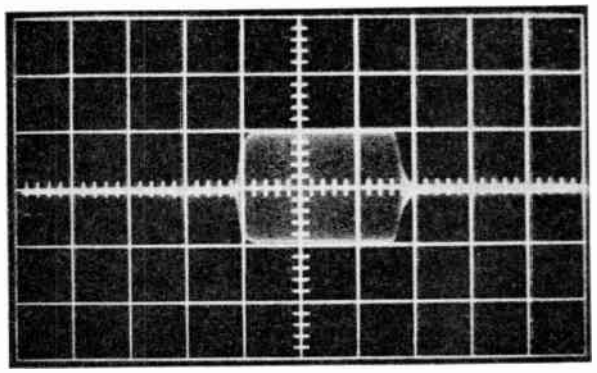


Fig. 5. 30-Mc/s input pulse. 1 division = 1 μs.

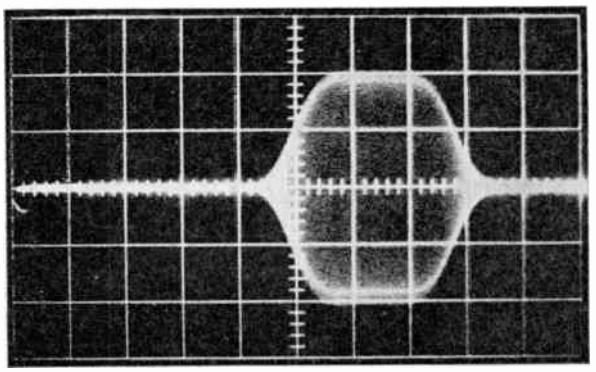


Fig. 6. 10-Mc/s output pulse. 1 division = 1 μs.

Figure 4 shows the measured and calculated amplifier frequency response; it will be seen that these are in good agreement, the slightly lower bandwidth of the measured amplifier probably arising from the 10-Mc/s output tuned circuit, which is ignored in the analysis. Figures 5 and 6 are photographic records of the input and output of the receiver when a 3-μs r.f. pulse is applied; the rise time of the pulse at the input is about 250 ns, while that of the output pulse is rather more than a microsecond, again in good agreement with calculations. The overall recovery time when used in a spin-echo system with a 1·5-μs pulse of 750 watts peak power is about 6 μs from the start of the pulse; the isolation between transmitter and receiver in this case was about 30 dB at the centre frequency, falling to 10 dB at a megacycle on either side.

5. Design of Second Detector

Measurement of relaxation time-constants from the receiver output presupposes a linear system, and in order to achieve linear rectification at the second detector it is necessary to drive it with an amplifier capable of delivering up to 25 V r.m.s. This gives a linear detection range of the order of 30 : 1, down to signals of some three-quarters of a volt. Below this level appreciable non-linearity occurs due to the curvature in diode characteristics. The intermediate frequency is accordingly chosen as being sufficiently low to enable the necessary drive voltage to be obtained fairly easily, yet not so low that signals from short-decay samples are distorted in rectification.

A further advantage of a 10-Mc/s i.f. is the relative ease of designing a phase detector at this frequency. The use of a phase detector is essential for samples whose signals are so weak that they are less than two or three times noise level. Contrary to what one might expect, linear detection is impossible under these conditions even if the combined level of signal plus noise is sufficiently large for the detector to be regarded as having an ideal characteristic.

A theoretical study of this situation has been made by R. E. Burgess,[7] who shows that the output of an ideal, linear detector in the presence of signal and noise is given by the formula

$$\bar{u} = (\pi/2)^{\frac{1}{2}} E \, {}_1F_1(-\tfrac{1}{2}, 1, -x^2) \qquad \text{.....(30)}$$

where $\bar{u}$ is the mean rectified voltage,

E is the r.m.s. value of the noise voltage,

x is the r.m.s. signal/noise ratio,

${}_1F_1$ is the hypergeometric function defined by

$${}_1F_1(-n/2,1,-x^2) = 1 + \frac{(-n/2)}{1!}\cdot\frac{(-x^2)}{1!} + \frac{(-n/2)(1-n/2)}{2!}\cdot\frac{(-x^2)^2}{2!} + \text{etc.} \qquad \text{.....(31)}$$

so that

$${}_1F_1(-\tfrac{1}{2},1,-x^2) = 1 + \frac{x^2}{2} - \frac{x^4}{16} + \frac{x^6}{96} - \text{etc.}$$

For large values of x it may be shown that

$${}_1F_1(-n/2,1,-x^2) = \frac{x^n}{\Gamma(1-n/2)}\left(1 + \frac{n^2}{4x^2} + \ldots\right)$$

whence

$${}_1F_1(-\tfrac{1}{2},1,-x^2) = \frac{x}{\Gamma(\frac{3}{2})}\left(1 + \frac{1}{4x^2} + \ldots\right) = 2\pi^{-\frac{1}{2}}x(1+\tfrac{1}{4}x^2+\ldots)$$

Thus, for small x,

$$\bar{u} \simeq (\pi/2)^{\frac{1}{2}} E(1+x^2)^{\frac{1}{2}} \qquad \text{.....(32)}$$

and for large x,

$$\bar{u} \simeq 2^{\frac{1}{2}} Ex \qquad \text{.....(33)}$$

showing that for large signals the detector is linear, whereas for small signals its output is proportional to the r.m.s. value of the mixture of signal and noise. It is evident from eqn. (30) that the change from square-law to linear behaviour depends only on x, the signal/noise ratio, and is independent of E, the absolute value of the noise voltage. Thus, linear

detection of small signals cannot be obtained simply by increasing the receiver gain. However, in a phase detector a reference signal is added which is phased coherently with respect to the signal being observed, and in this case Burgess' analysis no longer applies. Linear detection is now obtained provided the combined signal plus reference signal is large compared to the noise level; this can always be achieved by providing a sufficiently large reference signal.

It is the writer's intention to describe a phase-detection system at a later date. Meanwhile, it should be noted that for satisfactory operation a very high degree of stability is needed for the reference oscillator and the magnetic field which governs the n.m.r. frequency. These must either be linked to one another, or have individual short-term stabilities of a few parts in 10^9 if decays of several seconds duration, such as may occur in highly pure liquids, are to be studied.

6. Acknowledgments

The writer wishes to acknowledge the help of Mr. G. F. Miller, of the Mathematics Division of this Laboratory, in deriving formulæ for evaluating the Laplace transform integrals, and also for assistance in writing the Autocode programs. Acknowledgment is also due to Miss Brenda Webber, also of Mathematics Division, who supervised the machine calculations, and to Mr. M. J. Bonnett, who built the amplifier and carried out the measurements of frequency response.

The work described in this paper forms part of the research programme of the Basic Physics Division of the National Physical Laboratory, and is published by permission of the Director of the Laboratory.

7. References

1. J. Schwarz, "Spin-echo apparatus", *Rev. Sci. Instrum.*, **28**, pp. 780–9, October 1957.
2. J. C. Buchta, H. S. Gutowsky and D. Woessner, "Nuclear resonance pulse apparatus", *Rev. Sci. Instrum.*, **29**, pp. 55–60, January 1958.
3. K. Luszczynski and J. G. Powles, "Nuclear spin pulse apparatus", *J. Sci. Instrum.*, **36**, pp. 57–63, February 1959.
4. G. E. Valley and H. Wallman, "Vacuum Tube Amplifiers". (McGraw-Hill, New York, 1948.)
5. J. M. Pettit and M. M. McWhorter, "Electronic Amplifier Circuits". (McGraw-Hill, New York, 1961.)
6. F. F. Kuo, "Network Analysis and Synthesis". (John Wiley, New York, 1962.)
7. R. E. Burgess, "The rectification and observation of signals in the presence of noise", *Phil. Mag.*, **42**, p. 475, May 1951.

Manuscript received by the Institution on 23rd January 1964. (Paper No. 914.)

U.D.C. 621.373: 621.376

Highly-Efficient Generation of a Specified Harmonic or Sub-harmonic by Means of Switches

By

Professor
D. G. TUCKER, D.Sc.
(*Member*) †

Presented at a meeting of the Institution in London on 10th March 1964.

Summary: It is shown that a given harmonic can be generated efficiently by using a circuit in which the tuning is such that only the fundamental current flows in the input and only the desired harmonic voltage can appear at the output, and in which a balanced switching system is operated at the fundamental frequency by an external drive connected to the fundamental source. The switches cannot be replaced by rectifiers controlled by the voltages in the circuit except at the cost of dissipating a very large proportion of the available power in a d.c. load. The same principle of switching can be applied to make an efficient generator of sub-harmonics.

1. Introduction

The efficient generation of harmonics has been important for a long time, and many studies have been made in the past of methods of design of harmonic generators using rectifiers in order to obtain a particular harmonic with the greatest possible efficiency.[1] The efficiencies obtainable have, however, always been low, and it has fairly recently been shown by Page[2] that the generation of harmonics using non-linear resistance (e.g. rectifiers) can never give a power efficiency greater than $1/m^2$ for the harmonic of order m.

The advent of non-linear capacitance, usually in the form of the varactor diode, has recently led to a great deal of development of harmonic generators using such non-linear elements.[3] Since ideally such an element can absorb no power (although in practice the inevitable loss resistance to some extent limits the effect) it follows that if the circuit is so tuned that currents can flow (or voltages exist) only at the fundamental and a particular desired harmonic frequency, then all the power absorbed into the circuit from the source must, in principle, be passed into the harmonic-frequency load. Since it can be shown that the circuit can be conjugately-matched to the source, then it is clear that, in principle, all the *available* power from the source may be converted into a particular harmonic. Such harmonic generators have become important in microwave engineering, and power efficiencies in the range 50–70% are commonly obtained for low-order harmonics.

The obvious question then arises: if in a harmonic generator using rectifiers, the rectifiers are made to have a very low forward resistance and a very high backward resistance, so that very little power can actually be dissipated in them, why cannot practically all the available source power be converted into harmonic power? The answer is that the rectifiers are necessarily controlled by the voltage which appears across them. This consists not of the source voltage alone, but of the source voltage and the harmonic voltage together. Proper switching at the source frequency can be obtained only if a d.c. load is provided in addition to the harmonic load; and the d.c. load absorbs a major portion of the power. This is illustrated in a later section.

One possibility remains, however. If the source is caused to switch itself, not by the voltage appearing in the main circuit, but by driving the switches in synchronism with the source by means either of an electronic switch with a third electrode, or of a separate motor or solenoid drive receiving its power direct from the source, then the harmonic voltage cannot interfere with the switching, no d.c. load is needed, and all the source power can be converted into harmonic. In principle the driving mechanism for the switches absorbs no power, since ideally no work is needed to operate switches. The only loss is that due to current drawn by the switching electrode, say the grid in a valve or the base in a transistor, or due to inefficiency in the driving mechanism, and this can in many circumstances be kept to a very small percentage of the available source power. This is the method of harmonic generation to be discussed in detail in this paper. An additional source of loss is, of course, the tuned circuits, which cannot be completely free of resistance; but this applies to any kind of harmonic generator. Also the switches cannot be free of resistance.

It is appreciated that if mechanical switches have to be used, the method is restricted to fundamental frequencies below, say, 400 c/s. But it is thought that

† Department of Electronic and Electrical Engineering, University of Birmingham.

electronic switches may be found quite satisfactory, and that some may become available which would permit operation at very high frequencies (see Section 9).†

It is perhaps necessary to emphasize that it is the conversion of the *available* source power to a harmonic frequency which is one of the specific problems in electronic engineering and this is considered here. In power engineering, using say the electricity mains as the source, this consideration is quite irrelevant, and any harmonic generator which succeeded in transferring power to the output circuit and dissipated no power in the switches (or rectifiers) or in unwanted harmonic frequencies or d.c., would be considered 100% efficient. But to convert the *available* power means that the harmonic generator must be matched to the source, and this is an added complication.

It is of interest that in power engineering, efficient harmonic generators can be made on a completely different principle.[4] Three-phase circuits (or polyphase circuits of an appropriately higher order) are used with saturable inductors to synthesize a waveform, piece-by-piece, to contain dominantly the desired harmonic. It is quite possible that these methods would have advantages at much higher frequencies in certain circumstances.

Turning now to sub-harmonic generation, this has also been a matter of importance for some time, one of the most useful methods of achieving it being the regenerative modulator.[6, 7, 9] More recently non-linear capacitance has been used. The basic principle of these methods is essentially that an oscillatory circuit is set up at the sub-harmonic frequency with a non-linear element giving control at the input frequency. Since a circuit with non-linear capacitance (or inductance) driven at a 'pump' frequency is able to give, under suitable conditions, a negative resistance at a particular pair of terminals tuned to the sub-harmonic frequency, the requirement for sub-harmonic generation is easily achieved, and since the capacitance, if ideal, can absorb no power, the conversion is 100% efficient as in the non-linear capacitance type of harmonic generator.

A non-linear resistance circuit can give a negative resistance, however, only by means of feedback, and the regenerative modulator contains a very obvious feedback loop, since the output at the sub-harmonic frequency is fed back, through a harmonic generator and amplifier as appropriate, to form the modulating function of a modulator to which the input is applied. Such a device is very inefficient in the same way as the rectifier harmonic generator itself.

It is shown in Section 5 of this paper that the circuit using externally-driven switches, this time driven by the sub-harmonic output signal, forms a sub-harmonic generator ideally of 100% efficiency.

2. The Harmonic Generator

The system to be described is based on a ring modulator in which the input circuit is so tuned that the only current which can flow in it is at the source frequency ω_p, and the output circuit is so tuned that the only voltage which can exist in it is at the desired harmonic frequency $m\omega_p$. This tuning is, of course, an idealization; practical circuits can only approximate to it. The rectifiers of the usual ring modulator are replaced by four switches, as shown in Fig. 1, where mechanical switches are shown for clarity.

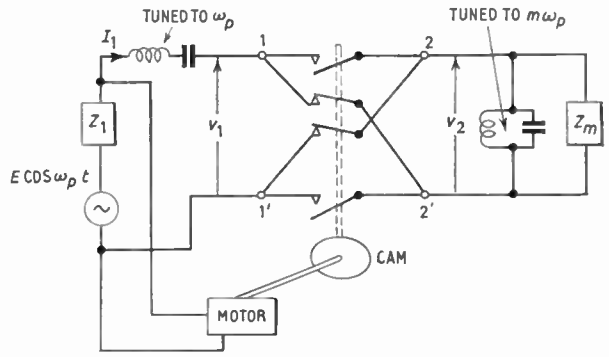


Fig. 1. Arrangement for generating even-order harmonics ideally with 100% power transfer.

These provide a commutating action under the influence of a motor-and-cam drive, the motor being driven synchronously by the source, and the phase of the cam being such that the switches open and close at prescribed instants relative to the zeros of the source waveform.

Now for the ring modulator tuned in the manner described above, but with an arbitrary switching frequency and the output tuned to a wanted sideband, Belevitch[5] has shown that a unity conversion-power-ratio is obtainable when the source and output load resistances are properly related. In the present case, his method of analysis can be adapted as follows.

Any reactance in Z_1 and Z_m may be tuned out (or absorbed in the tuned circuits), so it will be assumed that Z_1 and Z_m are pure resistances R_1 and R_m.

The current flowing into terminal 1 is a pure sinusoid, say, $I_1 \cos(\omega_p t + \theta)$. The switches operate in synchronism with this current, but with a relative phase θ rad, giving a square-wave switching function

† The need for a balanced circuit and for four switches is obviously a limitation on the application of the method at high frequencies; but an unbalanced form using only one switch and a mark/space ratio which is not unity may be derived from modulator theory.[12]

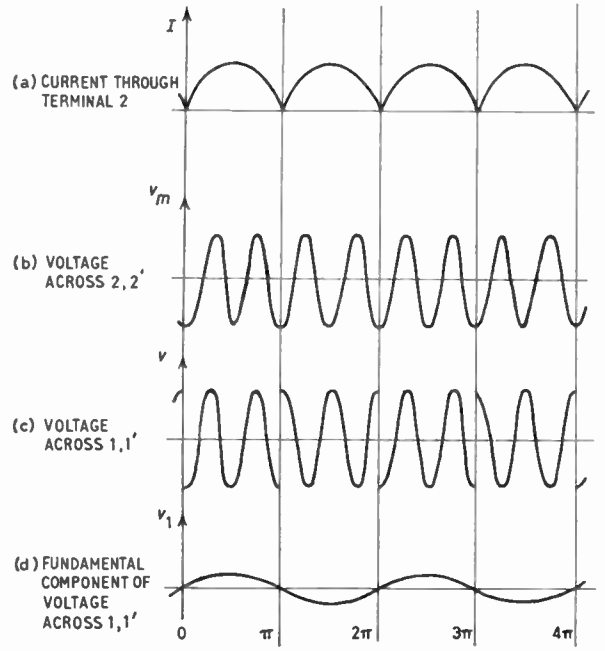


Fig. 2. Waveforms in harmonic generator of Fig. 1 for $m = 4$ and $\theta = 0$.

such that the current flowing out of terminal 2 is

$$I_1 \cos(\omega_p t+\theta) \cdot \frac{4}{\pi} \sum_{n=1,3,5\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n} \cos n\omega_p t \qquad \ldots\ldots(1)$$

This waveform is shown in Fig. 2(*a*) for $\theta = 0$ and $m = 4$. Now the only voltage which can exist across the terminals 2, 2′ is $v_m \cos(m\omega_p t+\phi)$, the voltage at frequency $m\omega_p$, as shown in Fig. 2(*b*). Therefore

$$v_m \cos(m\omega_p t+\phi) = (-1)^{(m-2)/2} \cdot \frac{2}{\pi} I_1 R_m \times \times \left[\frac{1}{m-1} \cos(m\omega_p t+\theta) - \frac{1}{m+1} \cos(m\omega_p t-\theta)\right] \qquad \ldots\ldots(2)$$

where it has been assumed that $(m-1)$ and $(m+1)$ are odd numbers, so that the harmonic generated, $m\omega_p$, must be of even-order.

Now for maximum power conversion, there must be a conjugate match at 1, 1′. The input impedance of the network (at frequency ω_p) to the right of 1, 1′ is the vector ratio of v_1/I_1, and $v_1 \cos(\omega_p t+\beta)$ is the component at frequency ω_p in

$$v_m \cos(m\omega_p t+\phi) \cdot \frac{4}{\pi} \sum_{n=1,3,5\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n} \cos n\omega_p t \qquad \ldots\ldots(3)$$

i.e. $$v_1 \cos(\omega_p t+\beta) = \frac{8I_1R_m}{\pi^2(m^2-1)^2} \times \times [(m^2+1)\cos(\omega_p t+\theta) - (m^2-1)\cos(\omega_p t-\theta)]$$

$$= \frac{16I_1R_m}{\pi^2(m^2-1)^2} \times \times (-m^2 \sin\theta \sin\omega_p t + \cos\theta \cos\omega_p t) \qquad \ldots\ldots(4)$$

The waveform corresponding to eqn. (3) is shown in Fig. 2(*c*), again for $\theta = 0$, and the corresponding component $v_1 \cos\omega_p t$ in Fig. 2(*d*).

It is clear that the input impedance at 1, 1′ is only a pure resistance when θ is either zero or a multiple of $\pi/2$ rad. When $\theta = 0$ or a multiple of π rad the input resistance is

$$\frac{v_1}{I_1} = \frac{16R_m}{\pi^2(m^2-1)^2} \qquad \ldots\ldots(5a)$$

and when $\theta = \pi/2$ or an odd multiple of $\pi/2$ rad, it is

$$\frac{v_1}{I_1} = \frac{16m^2R_m}{\pi^2(m^2-1)^2} \qquad \ldots\ldots(5b)$$

For intermediate values of θ, there is a complex input impedance. For a proper match, therefore, we require θ to be zero or a multiple of $\pi/2$ rad, and assuming then that R_1 is given, the optimum load resistance is, for $\theta = 0$ or a multiple of π rad,

$$R_{m(\text{opt})} = \frac{\pi^2(m^2-1)^2}{16} R_1 \qquad \ldots\ldots(6a)$$

and for $\theta =$ odd multiple of $\pi/2$ rad,

$$R_{m(\text{opt})} = \frac{\pi^2(m^2-1)^2}{16m^2} R_1 \qquad \ldots\ldots(6b)$$

Clearly the latter gives the more convenient values in practice. When a match is obtained, the input power is $E^2/8R_1$ and the output power is

$$\frac{v_m^2}{2R_m} = \frac{E^2}{8R_1} \qquad \ldots\ldots(7)$$

on substituting eqns. (2) and (6) and the appropriate values of θ. In other words, all power available from the source has been converted to power at the mth harmonic.

Although the circuit as analysed above can give only even harmonics, it is clear that if the switches are not operated simultaneously, but in such a way that even harmonics are developed in the switching function of the ring, then an odd harmonic can be obtained at the output terminals. Optimum matching conditions can be found for the odd harmonic when the angle of switching of each switch is specified.

The dual of the circuit given above, in which only the fundamental voltage exists at the input, and only the desired harmonic current flows in the output, can obviously be used as an alternative arrangement, and can be analysed in exactly the same way.

3. The Inefficiency of Self-switching

It has already been mentioned in the Introduction that if the external switching arrangements are

omitted, and the switches replaced by efficient rectifiers which switch under the action of the voltage appearing across them, then proper switching is obtained only if a suitable d.c. load is provided. Although a thorough investigation of the design and operation of such a circuit has not yet been made by the author, nevertheless the circuit of Fig. 3 has been set up and the effect of the d.c. load ($R_{d.c.}$) on the switching has been examined.

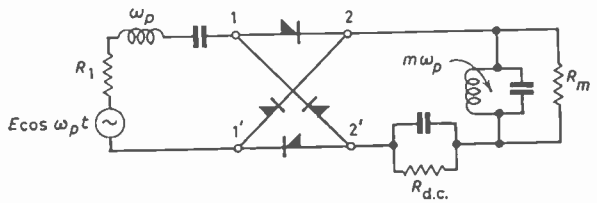


Fig. 3. Harmonic generator using self-switching.

The experimental circuit used germanium diodes type OA5, an input frequency of 5 kc/s, a harmonic frequency of 20 kc/s (i.e. $m = 4$), and $R_m = 90$ kΩ. The switching of the rectifiers, as would be expected, was not unduly dependent on the source resistance R_1; if the series tuned circuit were ideal, it would presumably not be dependent on it at all. But it should be stated that the experimental observations were made with a very low source resistance of only about 5 ohms. The easiest way to observe how the switching is operating is to examine the voltage waveform across terminals 1, 1′, and to compare it with that of Fig. 2(*c*) for the externally-switched circuit. With no d.c. load (i.e. $R_{d.c.} = 0$) the waveform was that shown by the oscillogram of Fig. 4(*a*). (In all these oscillograms the input e.m.f. waveform is also shown so that the phase of switching may be seen.) It is observed that the rectifiers permit transmission in the circuit only for a small part of the cycle. Consideration of the detailed operation of the rectifiers shows that all four are conducting for part of the cycle, thus draining much power uselessly from the source. When the d.c. load is made optimum the waveform is as shown in Fig. 4(*b*), and it is seen that the rectifiers are operating almost correctly as commutating switches, but that the waveform now contains, in addition to the switched fourth harmonic (as Fig. 2(*c*)) a component of switched d.c. which just equals the peak amplitude of the harmonic. When the d.c. load is increased to about twice the optimum the waveform shown in Fig. 4(*c*) is obtained. In this waveform it is evident that the d.c. load is excessive, and although the switching is almost correct, too much of the power is being diverted into the d.c. output.

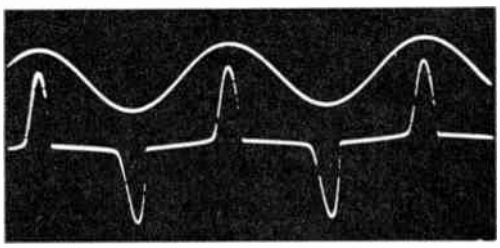
It is thought that these results illustrate why a d.c. load is necessary in the self-switched case, and why this circuit cannot be efficient. It can easily be shown that the best efficiency, assuming ideal rectifiers, is in accordance with Page's law, i.e. $1/m^2$.

We thus conclude that the method of Section 2 is the only way of 'getting round' Page's law.

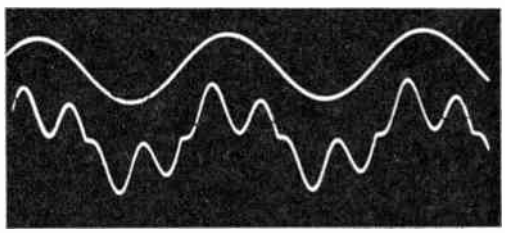
4. Experimental Results with the Efficient Harmonic Generator

For experimental investigation of the mechanism of harmonic generation, it was decided to use mechanical switches since they give very low resistance and readily-adjustable phase of operation. It is unlikely that the power absorbed by a motor driving the switches could be small enough to permit really efficient harmonic generation at power levels less than a few watts, but the use of high-speed relays can give good results at much lower power levels. For most reliable and consistent operation, however, a rotary switch appears to be best.

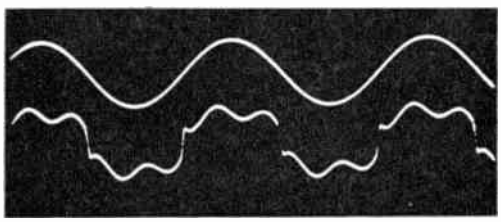
The rotary switch method uses a drum attached to the motor shaft with conducting segments and switch contacts. Since there is no movement (ideally) of the switch contacts, contact bounce is minimized. But, nevertheless, it requires much care to obtain good switching over any long period of time at low applied



(*a*) $R_{d.c.} = 0$.



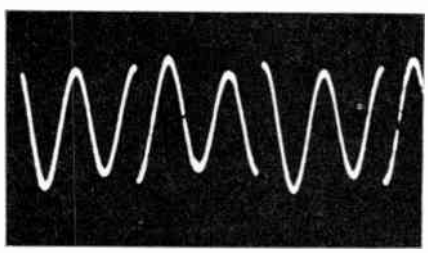
(*b*) $R_{d.c.} = 42$ kΩ.



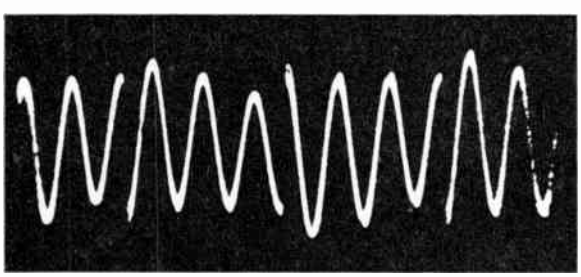
(*c*) $R_{d.c.} = 92$ kΩ.

Fig. 4. Waveforms at terminals 1, 1′ of Fig. 3. Upper waveform in each is applied e.m.f.

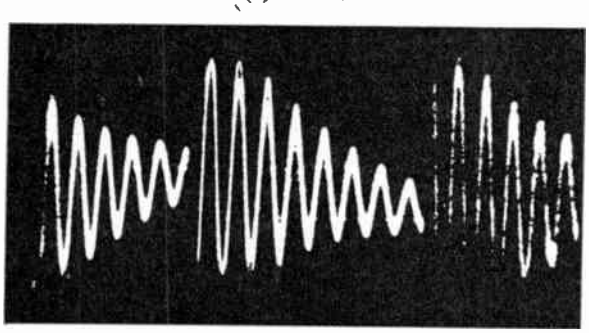
voltages. The speed of rotation of the drum can be reduced by using a motor or gearing such that the angular speed is an integral fraction of the angular frequency from which the harmonic is to be generated, and having a series of conducting segments equally spaced round the drum. This method was used by the author in an experimental check of the theory, and his apparatus gave quite good results for a 400 c/s fundamental frequency. It is more difficult to get the switching phases correct by this method, but the oscillograms of the voltage waveform at terminals 1, 1′ shown in Fig. 5 (for various values of m) show only small errors. (They should be compared with the ideal waveform of Fig. 2(*c*).) It is clear from the experimental results that the circuit operates as described in Section 2, and the required harmonic can be generated without the need for any d.c. load.



(*a*) $m = 4$.



(*b*) $m = 6$.



(*c*) $m = 16$.

Fig. 5. Waveforms at terminals 1, 1′ of Fig. 1 with rotary switch. Note that the horizontal scale is not the same in these oscillograms: also that some contact trouble is evident in (*c*), where successive traces of the oscilloscope have not exactly coincided due to varying contact resistance. For all three values of m, $R_1 = 15\Omega$ and $R_m = 50\ \text{k}\Omega$ and the vertical scale of the oscillograms is the same.

It is evident from the oscillogram for $m = 16$ that the Q-value of the output tuned circuit is not really high enough. It was actually effectively only about 30; and if we think about the functioning of the circuit in terms of each switching operation exciting a transient oscillation in the tuned circuit, then clearly the decrement of this oscillation (or 'ringing') is too great. The detailed design of the harmonic generator requires further study if really high efficiency and purity are to be attained. But it is obvious that a completely pure harmonic can never be obtained with a single tuned circuit in the output, since the load resistance ensures that the Q-value cannot be infinite. The assumption of a pure output waveform in the analysis of Section 2 is to this extent unjustified and must be regarded as an approximation.

The efficiencies obtained in these experiments (leaving out of account the power absorbed in the motor) were in the range 25–50% for matched impedances; the bulk of the losses must evidently occur in or because of the tuned circuits, and these losses would occur to a similar extent in the rectifier harmonic-generator of Fig. 3. Thus, assuming that the power absorbed by the switch drive does not exceed, say, 25% of the source power, it becomes clear that the switch circuit gives a really enormous improvement on the higher harmonics. For example, with $m = 16$, the switch circuit gives an efficiency of about 23% on this basis (30% not counting the motor loss), but the rectifier circuit gives only about $30/m^2$, i.e. 0·12%.

5. The Efficient Sub-harmonic Generator

Since the harmonic generator described in Sections 2 and 4 is a linear device, it obeys the concept of reciprocity in the sense that if the e.m.f. at ω_p is removed from the left-hand side in Fig. 1, and a current source at $m\omega_p$ is applied at the right-hand side, then a current at frequency ω_p is generated in the left-hand circuit. This assumes, of course, that the switches continue to be driven by the signal at ω_p, and so presupposes the existence of this signal before it can, strictly speaking, start to be generated. This means, in practical terms, that the sub-harmonic generator is not self-starting. However, except for $m = 2$, the regenerative modulator type of sub-harmonic generator is not naturally self-starting either. It is likely that the switch type of circuit would be harder to start than the regenerative modulator, but this has not yet been investigated.

The dual arrangement, as shown in Fig. 6, which uses an e.m.f. source at $m\omega_p$ instead of a current source, would probably be more convenient in practice. The analysis of this, assuming for simplicity that the switching is in phase with the output, is as follows:

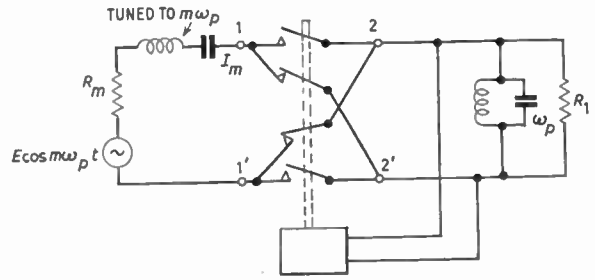


Fig. 6. Sub-harmonic generator.

The input current at terminal 1 is a sinusoid $I_m \cos m\omega_p t$, as shown in Fig. 7(*a*). The current at terminal 2 is therefore

$$I_m \cos m\omega_p t \cdot \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n} \cos n\omega_p t \qquad \ldots\ldots(8)$$

The waveform of this is shown in Fig. 7(*b*) for $m = 4$. Now the only voltage which can exist across terminals 2, 2′ is v_1 at frequency ω_p as shown in Fig. 7(*c*). This is therefore

$$v_1 = (-1)^{(m-2)/2} \cdot \frac{2}{\pi} I_m R_1 \left[\frac{1}{m-1} - \frac{1}{m+1}\right]$$

$$= (-1)^{(m-2)/2} \cdot \frac{4}{\pi} I_m R_1/(m^2-1) \qquad \ldots\ldots(9)$$

The voltage across terminals 1, 1′ is therefore

$$v_1 \cos \omega_p t \cdot \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n} \cos n\omega_p t \qquad \ldots\ldots(10)$$

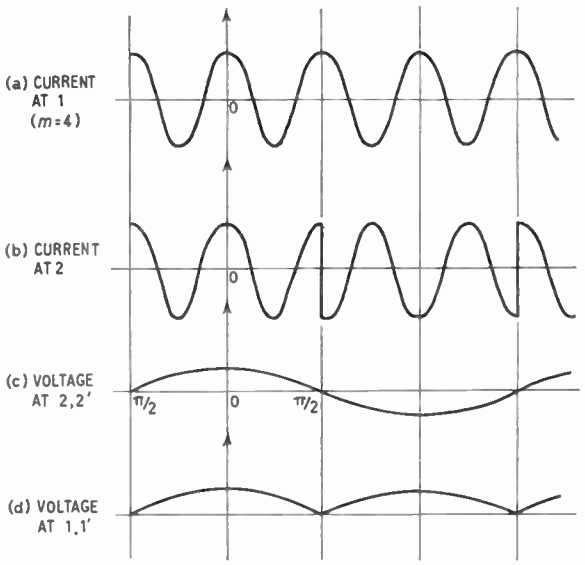


Fig. 7. Waveforms in sub-harmonic generator of Fig. 6 ($m = 4$).

the waveform of which is shown in Fig. 7(*d*). The voltage at frequency $m\omega_p$ is therefore the component (v_m) at $m\omega_p$ in this voltage wave, and the input resistance is

$$\frac{v_m}{I_m} = \frac{16}{\pi^2} \cdot \frac{R_1}{(m^2-1)^2} \qquad \ldots\ldots(11)$$

If therefore R_m is made equal to this value, a match is obtained, and efficient sub-harmonic generation is achieved. In the above analysis, m is, of course, even.

6. Conclusions

It has been shown that a given harmonic can be generated efficiently (at least in principle) by using a harmonic-generator circuit in which switches (mechanical or electronic) are driven at the fundamental frequency by an external drive connected to the fundamental source. It has also been shown that it is not possible to replace the switches by rectifiers 'switched' by the voltages in the circuit except at the cost of dissipating a very large proportion of the available power in a d.c. load; in these circumstances the power efficiency of generation of the mth harmonic cannot exceed $1/m^2$ as given by Page's theory, even if the rectifiers are ideal.

Sub-harmonics may be generated efficiently by the use of essentially the same system, although it may be difficult to arrange for self-starting.

7. Acknowledgments

The author is not quite sure whether the work described in this paper is new or not; but if it is, then a large share of the credit is due to Professor V. Belevitch of the University of Louvain, Belgium, who first interested the author in the matter. The author's thanks are due to Mr. B. L. J. Kulesza and Mr. C. Williams (Research Associate and Chief Technician respectively in the Department of Electronic and Electrical Engineering at the University of Birmingham) for the experimental work reported in Section 4, and to Mr. T. W. Stasiw (Athlone Fellow attending the postgraduate course in Information Engineering at the University of Birmingham) for that reported in Section 3.

8. References

1. See, e.g. V. Belevitch, "Théorie des Circuits Non-Linéaires en Régime Alternatif", Chap. 8 (Librairie Universitaire Uystpruyst, Louvain, 1959).
2. C. H. Page, "Frequency conversion with positive non-linear resistors", *J. Res. Nat. Bur. Stand.*, **56**, p. 179, 1956.
3. See, e.g. L. A. Blackwell and K. L. Kotzebue, "Semiconductor–Diode Parametric Amplifiers", Chap. 4.4 (Prentice-Hall, Englewood Cliffs, N.J., 1961).
4. L. R. Blake, "The frequency tripler", *Proc. Instn Elect. Engrs*, **100**, Part II, p. 296, 1953.
5. V. Belevitch, *loc. cit.*, Chap. 3, Section 3.

6. R. L. Miller, "Fractional frequency generators utilizing regenerative modulation", *Proc. Inst. Radio Engrs*, 27, p. 446, 1939.
7. D. G. Tucker and H. J. Marchant, "Frequency division without free oscillation", *P.O. Elect. Engrs J.*, 35, p. 62, 1942.
8. J. R. Cannon, "The regenerative modulator", *A.T.E. J.*, 10, p. 159, 1954.
9. G. H. Parks, "Symmetrical transistors", *J.Brit.I.R.E.*, 21, p. 79, January 1961.
10. G. T. Wright, "Space-charge-limited solid-state devices", *Proc. Inst. Elect. Electronics Engrs*, 51, p. 1642, November 1963.
11. S. Brojdo, "Characteristics of the dielectric diode and triode at very high frequencies", *Solid-State Electronics*, 6, p. 611, 1963.
12. D. P. Howson, "Single-balanced rectifier modulators: an analysis which includes the effect of changing the mark/space ratio of the switching signal", *Proc. Inst. Elect. Engrs*, 109C, p. 357, 1962 (I.E.E. Monograph No. 500E, January 1962).

9. Appendix. Possible Electronic Switch Arrangements

As mentioned in the Introduction, suitable electronic switch arrangements may be found to permit the operation of the harmonic or sub-harmonic generator at high and very high frequencies. Over a very useful range of frequencies a double-triode arrangement as shown in Fig. 8 could be used. Existing types of thermionic triode use a considerable power in the heaters, and if this had to be taken into account in calculating the efficiency of a harmonic generator, then the efficiency could not be very high. But as this heater power does not come from the signal source (except in applications to a power-system harmonic generator, where perhaps silicon-controlled rectifiers could be used instead) it can probably justifiably be neglected. Even so, the relatively high resistance of ordinary triodes will lead either to reduced efficiency or to rather high circuit impedances, and the circuit voltages need to be high.

Transistors can be used as switches in this application, and have the big advantage of needing only low voltages for proper operation. The simplest device is the symmetrical transistor.[9] Although in all transistor arrangements the current drawn from the switching control circuit is likely to be a larger proportion of the current in the main circuit than in the case of thermionic valves, the power loss due to this cause is probably negligible.

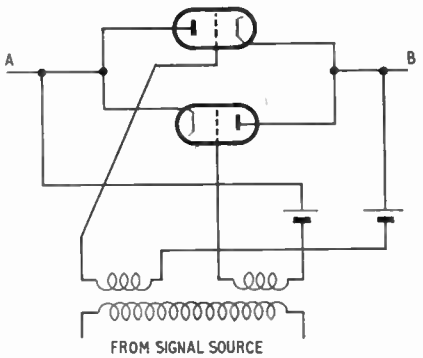


Fig. 8. Triode switch arrangement. The circuit between A and B is opened and closed by the controlling signal reasonably independently of the voltage between A and B.

The surface-channel triode,[10] which has already been successfully demonstrated, appears to be suitable for this application at frequencies up to about 100 Mc/s.

At very high frequencies, including the usual microwave range, the harmonic and sub-harmonic generators described here have potentialities and may be competitive with circuits using varactors if the prospective development of the planar space-charge-limited dielectric triode[10, 11] proves successful.

Manuscript first received by the Institution on 27th January 1964 and in final form on 23rd April 1964. (Paper No. 915.)

DISCUSSION

Under the Chairmanship of Mr. B. F. Gray

Dr. F. J. Hyde: Professor Tucker has himself recognized the drawbacks of applying electronic switching in the harmonic generator which he has described. Faster electronic switches than those now available will undoubtedly be developed, but it is interesting to note that the varactor diode, which is already established as an efficient harmonic generator, can now be made† with a cut-off frequency in excess of 1000 Gc/s.

† C. A. Burrus, "Formed point-contact varactor diodes utilizing a thin epitaxial gallium arsenide layer" *Proc. Inst. Elect. Electronics Engrs*, 51, p. 1777, 1963.

Mr. D. L. Hedderly: I should like to make two points concerning Professor Tucker's paper. The first is to draw attention to the possibility of obtaining power from the d.c. bias battery in harmonic generator circuits; this possibility has been mentioned previously in connection with harmonic generators using varactor diodes.‡ One way in which this might be achieved is to use a charge multiplying device; that is one in which the charge storage

‡ A. Uhlir, "Applications of New and Recently Developed Diodes", Microwave Associates, Inc., 1959.

is more than complete so that more charge crosses back over the junction than was initially injected. Using such a device it is easy to show† that power can be drawn from the bias battery although whether or not it can be constrained to appear as a useful output is an open question. In this connection a charge multiplying diode has recently been described by Lindmayer and Wrigley.‡

The second point concerns the instantaneous power flow in harmonic generator circuits. Since the supply of power to the output filter in a harmonic generator is not steady, as suggested by the mathematics, but periodic (at twice the input frequency in the case of Professor Tucker's circuits), then it is essential that the energy stored in the output filter should decrease; otherwise no power would flow to the load. This periodic decrease and increase in the energy stored in the output filter inevitably means that the output will contain frequencies other than the wanted one.

Professor D. G. Tucker (*in reply*): Dr. Hyde has rightly pointed out what a large lead the varactor diode has over electronic switches in regard to operation at microwave frequencies. But it seems worthwhile to examine harmonic-generation circuits other than those in current use, and there is a strong possibility that, when the new solid-state triodes emerge, a development of my circuit in unbalanced form will give higher efficiencies than those obtainable from varactors.

Mr. Hedderly's first point is interesting, but I have no further information on the matter. His second point is an important one, but I do not think my paper is misleading in this respect. I included Fig. 5(*c*) especially to bring out this point. I expect it is the idealized tuning arrangements assumed in the analysis which have caused Mr. Hedderly to raise the matter; but if real tuned circuits with finite Q-factors are taken, the analysis becomes much more difficult and the essentially simple conception of the circuit becomes obscured. I agree with Mr. Hedderly that high efficiency and high purity are not consistent with one another, but this is a feature of all harmonic generators involving tuning and not just of my circuit.

† D. L. Hedderly, "An analysis of circuit for the generation of high-order harmonics using an ideal non-linear capacitor", *Trans.I.R.E.* (*Electron Devices*), **ED-9**, pp. 484–91, November 1962.

D. L. Hedderly and J. J. Sparkes, "A parametric negative resistance in transistors", *Proc. International Symposium on Semiconductor Devices*, Paris, February 1961, Editions Chiron, Paris.

‡ J. Lindmayer and C. Wrigley, "A new aspect of the semiconductor diode", *J. Electronic Control*, **14**, No. 3, pp. 289–301, March 1963.

This paper was one of four presented at a meeting on "Some New Possibilities in Parametric Devices." The other three papers together with associated discussion, were published in the June issue of *The Radio and Electronic Engineer* as follows:

K. L. Hughes and D. P. Howson, "Double-sideband Parametric Conversion Using Non-linear Resistance and Capacitance".

D. P. Howson and A. Szerlip, "Parametric Up-conversion by the use of Non-linear Resistance and Capacitance".

D. G. Tucker and K. L. Hughes, "Parametric Amplifiers and Converters with Pumped Inductance and Capacitance".

U.D.C. 621.397.334

An Appraisal of Some Decoding Systems for Three-gun N.T.S.C. Colour Television Receivers

By

D. R. BIRT†

AND

K. G. FREEMAN, B.Sc.
(*Associate Member*)‡

Summary: Results are given of some practical and theoretical investigations of several types of decoding systems for use with three-gun N.T.S.C. receivers. It is concluded that the $(R-Y)/(B-Y)$ type of decoder is the simplest and most convenient from the point of view of design and adjustment and is likely to be the least critical in operation.

1. Introduction

In domestic colour television receivers employing the three-gun shadow-mask tube and operating with the N.T.S.C. system of signal coding a variety of signal decoding methods are possible. Early receivers made use of the full bandwidth I/Q system of decoding followed by matrixing to obtain the required red, green and blue tube drive signals but, as is well known, this approach leads to a complicated design of receiver with a number of disadvantages.

If colour difference operation of the shadow-mask tube is employed—the cathodes of the guns being driven with the luminance signal and the grids with the $(R-Y)$, $(G-Y)$ and $(B-Y)$ colour difference signals—a better performance is possible, particularly on monochrome transmissions. Furthermore, by accepting only a small loss in colour resolution and resorting to equi-band operation a very much simpler chrominance signal decoder is possible, and over the years quite a number of equi-band decoders (such as $(R-Y)/(B-Y)$, the so-called X/Z decoder of R.C.A. and others[1, 2]) have been proposed and employed in domestic receivers in the U.S.A. However, although some of these decoders have been shown to have obvious advantages, there is very little evidence regarding their overall performance, particularly when the spreads associated with any component values, etc., are taken into account.

Although at the time of writing a European colour standard is not agreed upon, the U.K. receiver industry must, nevertheless, consider possible designs for domestic colour receivers for use with the N.T.S.C. system. In the absence at present of any alternative suitable display device such designs will be based upon the shadow-mask tube and must include amongst other things a chrominance signal decoder which would incorporate two or more synchronous detectors and a matrix circuit for deriving the colour difference signals.

In this paper a number of possible colour difference decoding systems, employing thermionic valves for the final matrixing, will be examined in some detail from both theoretical and practical points of view in order to assess their relative merits.

From a practical point of view the suitability of a particular design must be considered in relation to any other limitations imposed in the receiver, such as the preferred choice of h.t. voltage. The decoder should be easy to set up and, where necessary, to adjust, and it should be immune, for example, to the effects of h.t. line ripple. From the point of view of obtaining acceptable colour rendition in the final picture the design should be such that normal tolerance components may be used with confidence, or, where this is out of the question, simple adjustment must be possible to take account of initial spreads or the variation of parameters with receiver life. The question of providing correction for the use of tubes with all-sulphide phosphors with N.T.S.C. transmission primaries will also be dealt with.

The various systems of decoding to be considered in this paper determine mainly the method of colour difference signal matrixing employed and only place restrictions upon the synchronous demodulators in respect of gains and phases. From a theoretical standpoint, therefore, a detailed consideration of the various possible methods of synchronous demodulation is not essential for the present discussion. Unfortunately, a preliminary examination of the effects of non-linearities and of wide sampling angles in synchronous demodulators has, in any case, led to some formidable mathematical problems which so far remain unsolved.

In respect of circuit complexity there is probably not a great deal to choose between the various well

† Formerly with Mullard Research Laboratories, now with Redifon Ltd., Crawley, Sussex.

‡ Mullard Research Laboratories, Redhill, Surrey.

known methods of synchronous demodulation, which have been described in the literature,[2] and the choice of method is largely dictated by convenience and the level of signal required from the detectors.

The authors therefore consider that the omission of a detailed assessment of synchronous demodulators does not significantly affect the object of this paper, which is to compare a number of decoding *systems*. Nevertheless it is hoped that a separate analysis of demodulators may be possible at a later date.

2. Decoder Requirements

2.1. *Compatibility with the Remainder of the Receiver Circuits*

It will be obvious that any decoder matrix arrangement proposed must not add to the overall problems of the receiver designer by requiring conditions which are not easily met by the other circuits or which would add materially to the overall cost of the receiver. For example, it would hardly be sensible to design a matrix requiring say a 400-V h.t. line when the rest of the receiver was based upon one of 250 V. Clearly, there are many factors of this nature to be taken into account when attempting an integrated and economic receiver design but as these will depend upon the particular design requirements, this point will not be elaborated upon here.

2.2. *Equal vs. Unequal Tube Drive Operation*

As indicated in the introduction, it is now general practice to use colour difference operation of the shadow-mask tube, with the luminance signal applied to the cathodes and the colour difference signals to the grids. The colour difference signals, of course, vanish on monochrome transmissions and the grey scale rendition of the tube will then depend upon the beam current vs. drive characteristics of the guns, the relative phosphor efficiencies and the luminance drives on the cathodes. One simple arrangement is to put the same amplitude of luminance signal on all three cathodes and then adjust the tube slopes by means of the g_1 and g_2 potentials in order to obtain grey scale tracking, but in practice the widely different gun slopes necessary may make it difficult to match the shapes of the transfer characteristics sufficiently well to ensure a satisfactory grey scale.

It is therefore now usual to adjust the gun characteristics to be as nearly identical as possible and to supply the three cathodes with suitably unequal proportions of the luminance signal (unequal drive operation). Clearly for correct rendition the colour difference signals applied to the tube grids must be adjusted in the same proportions. For a typical phosphate tube such as the 21CYP22 the proportions of signal required by the red, green and blue guns to give Illuminant 'C' white are approximately in the ratio 1 : 0·8 : 0·6, whereas for the AX53-14 all-sulphide tube they are approximately in the ratio 1 : 0·85 : 0·7. The difference is only partly due to the different luminance contributions to white resulting from the different primary colour points and is principally the result of the different relative efficiencies of the two kinds of phosphor materials.

In either case, depending upon the tube operating conditions the appropriate drive requirements are between 70 and 100 V for the luminance signal and up to 200 V peak-to-peak for the colour difference signals.

2.3. *Adjustment for Differences in Tube Drive Requirements*

As stated above, differences in the relative drive requirements of the red, green and blue guns can occur from tube to tube of a particular type. Clearly, if correct grey scale rendering is to be obtained, provision must be made for adjusting the relative luminance drives, and this is normally done.

However, it must also follow that if correct colour rendering is to be obtained the colour difference signal drives must also be adjusted in the correct proportions.

The colour errors which arise in an N.T.S.C. receiver with a phosphate tube in which one of the guns requires a value of colour difference drive different from the nominal have been computed for a number of cases, assuming transmission and display gammas of 0·5 and 2 respectively. The resulting errors in chromaticity and luminance may be plotted on the 1931 C.I.E. chromaticity diagram.

Figures 1, 2 and 3 show the errors which occur in a phosphate display when the red, green and blue guns respectively require 20% more colour difference drive than the nominal amount. The corners of the triangle represent the primary phosphors of the display (assumed to be the same as the transmission primaries which in turn are assumed to be the standard N.T.S.C. set balanced to Illuminant 'C'). The dots indicate the chromaticities of a number of selected colours within the system gamut and the arrowheads the reproduced colours. The accompanying figures show the ratio of reproduced to intended luminance where significantly different from unity within the accuracy of the computations.

It will be seen that some of the colour shifts and luminance errors are hardly negligible. However, it should be borne in mind that a colour shift of a given distance on the diagram varies in its subjective perceptibility over the colour triangle and, for example, quite large errors of colour co-ordinates near green may pass unnoticed whereas co-ordinate errors near primary blue and red can be fairly critical. At present it is difficult to establish the full significance of colour

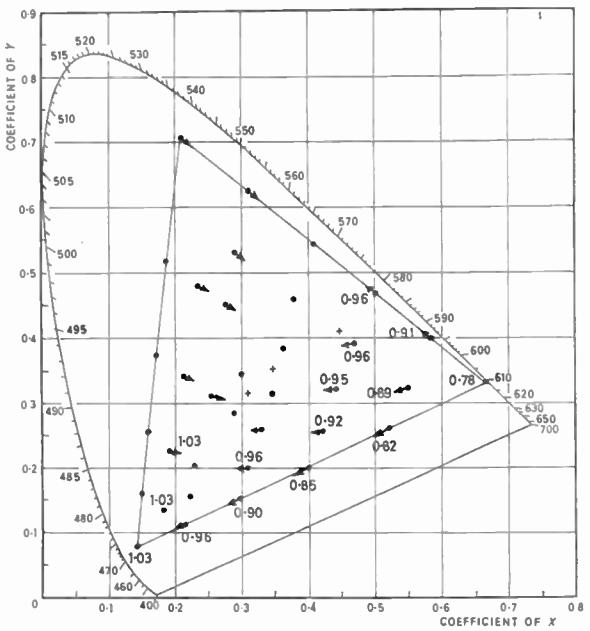


Fig. 1. Colour errors on phosphate tube requiring 20% more red colour difference drive than supplied.

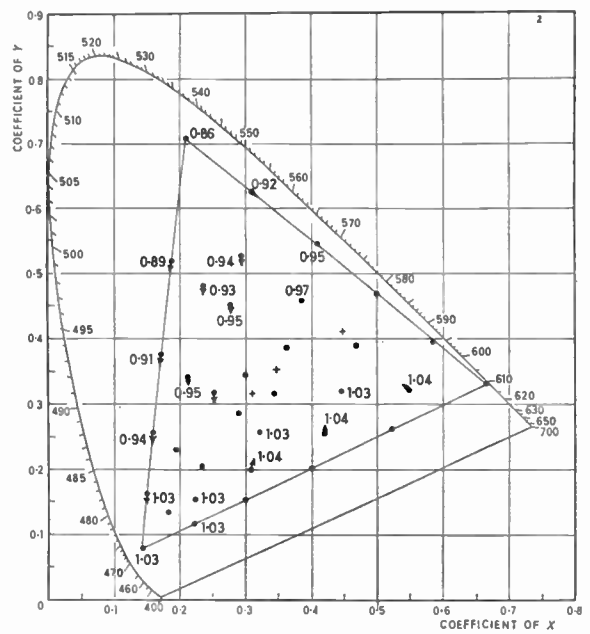


Fig. 2. Colour errors on phosphate tube requiring 20% more green colour difference drive than supplied.

error diagrams of the kind shown in Figs. 1, 2 and 3 and elsewhere in this paper. Subjective experiments are now under way to endeavour to correlate such error diagrams with the acceptability or otherwise of the colour pictures in which they result. Correlation of the practical results with those calculated is complicated by the fact that in practice the display

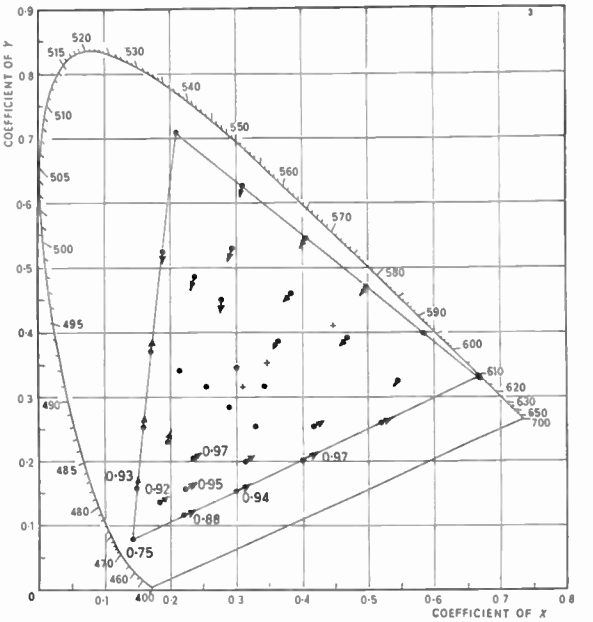


Fig. 3. Colour errors on phosphate tube requiring 20% more blue colour difference drive than supplied.

gamma is greater than 2 and this modifies the colour errors, some being reduced and others being increased as compared with those calculated. Furthermore, preliminary results from the experiments indicate that quite large errors in terms of C.I.E. co-ordinates may be subjectively tolerable and it is hoped to publish detailed results in due course. Nevertheless, calculated error diagrams of the type shown still make it possible to establish which parameters in a given system of decoding, and which system of decoding, are most critical.

In fact it is felt that errors of the magnitude shown in Figs. 1, 2 and 3 cannot safely be ignored completely and that where spreads in tube drive characteristics of this order can occur, provision should be made in the decoder for suitably adjusting the colour difference drive signals obtained from the decoder matrix. Such adjustments should, of course, be easy to effect. In this respect a number of possibilities are open to the designer. The usual practice hitherto has been to design the decoder to provide the correct proportions of output signals for a nominal tube (be it phosphate or sulphide). However, this may complicate the provision of adjustments to correct for the variations from the nominal drive requirements which may be encountered in a particular tube.

Another possibility is to design the decoder on the assumption that the tube requires equal drive and then to tap down the colour difference outputs from the matrix to suit the particular tube. Although this may result in a restriction of the maximum usable

output (see Section 3.1.1), it does allow the possibility of adjusting the colour difference taps to match those of the luminance drive adjustments. A ganged adjustment of luminance and colour difference drives could even be envisaged. This type of design will be termed 'unweighted'. Obviously, other design arrangements than these two are possible but they will not be considered in this paper.

2.4. *Correction for Sulphide Display Primaries*

It is now general practice to employ all-sulphide primary phosphors in shadow-mask tubes as these yield a brighter picture and eliminate the colour trailing on 'panned' scenes due to phosphor afterglow (in particular of the Willemite green) which occurs in the 'phosphate' tube. However, the sulphide primary phosphors are not all obtainable with the same colour points as specified for the transmission system and in particular the red and green phosphors are significantly desaturated.

The C.I.E. colour points of the primary phosphors of the 21CYP22 phosphate and the AX53-14 all-sulphide tubes are compared below.

	21CYP22		AX53-14	
	x	y	x	y
Red	0·67	0·33	0·64	0·34
Green	0·21	0·71	0·26	0·58
Blue	0·14	0·08	0·15	0·07

It will be apparent that the colour gamut of the sulphide tube is reduced as compared with that of the phosphate tube (and therefore of the transmission primaries normally used), so that, for example, the most saturated greens cannot be reproduced. However, it will also be evident that if no correction is made errors in reproduction will occur over most of the colour gamut.

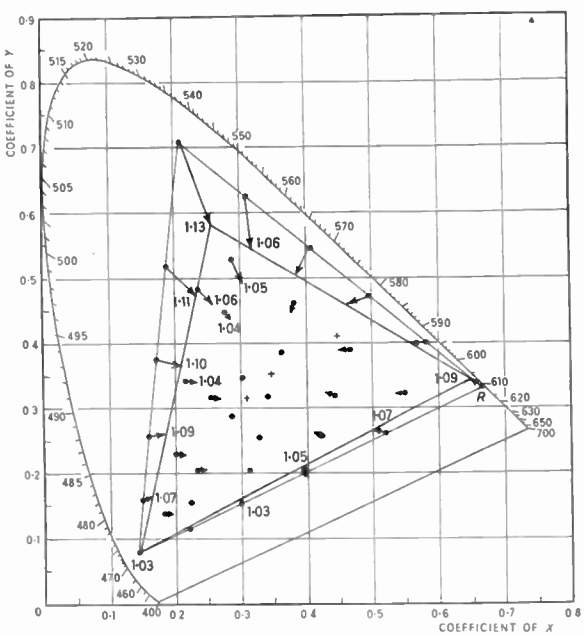


Fig. 4. Errors on sulphide tube after approximate colour point correction.

This problem has been discussed in more detail elsewhere[3, 4] and it has been shown that substantially adequate correction over most of the reproduction colour gamut can be achieved in the receiver by a slight modification of the colour difference drive outputs to the following form:

$$[R-Y] = 1{\cdot}155(E'_R - E'_Y)$$

$$[G-Y] = 1{\cdot}07(E'_G - E'_Y)$$

$$[B-Y] = 1{\cdot}06(E'_B - E'_Y)$$

(To a good approximation these output increases can be obtained by a 6 or 7% increase in chrominance signal gain together with a further increase in the $(R-Y)$ signal of 9 or 10%.)

The resulting errors with this arrangement are shown in Fig. 4 and are seen to be comparatively small. These increases in colour difference drive are of course over and above the adjustments in colour difference drives necessary to take account of the fact that the phosphor efficiencies of the sulphide tube differ from those of a phosphate tube, so that for a nominal AX53-14 sulphide tube, with $R : G : B$ drives required to be in the ratio 1 : 0·85 : 0·70, the relative colour difference signal drives required for the tube grids when colour point correction is used are given by

$$1{\cdot}155(E'_R - E'_Y), \quad 0{\cdot}91(E'_G - E'_Y) \quad \text{and} \quad 0{\cdot}74(E'_B - E'_Y)$$

respectively. Again it is desirable that these adjustments of colour difference drives should be easy to effect in the decoder matrix.

2.5. *Decoder Errors*

Quite apart from the errors arising as a result of incorrect colour difference signal drives to the tube it is apparent that colour errors can occur if any of the components or valve parameters differ from their design values—and with inevitable production spreads this is unavoidable. Furthermore, errors in reproduction can also result if the synchronous demodulators have an incorrect gain or phase. (The choice of a poor design of synchronous detector may also lead to crosstalk and other errors but these will not be considered in this paper.)

It is therefore necessary to investigate the colour errors which can occur due to the use of normal tolerance components (such as 10% resistors) or due to errors in the demodulation phases or gains and to establish whether these errors are tolerable or not.

If the errors are not acceptable, of course, provision for adjustment must be provided in the decoder. As will be shown, this may not be easy with some decoders.

2.6. *The D.C. Component*

From laboratory studies using RGB-driven monitors it has been established that the d.c. levels of the black levels of the three channels should be carefully maintained. It therefore follows that in a receiver employing colour difference operation of the display the d.c. components of the colour difference signals applied to the grids should be well maintained—ideally 100%. Failure to do this can result in an overall coloration of the picture which is a variable function of the picture content and which can be quite objectionable. It is therefore necessary either to design the decoder to have d.c. coupling from the synchronous detectors to the tube grids or to use some method of d.c. level re-insertion. Since the colour difference signals can be positive- or negative-going relative to the zero-level corresponding to monochrome pictures this latter alternative usually means the employment of some form of keyed clamping, probably operating during the line back-porch period.

2.7. *Other Requirements*

It is hardly necessary to point out that the colour difference decoder (and of course the luminance amplifier) should provide adequate levels of signal for driving the tube to its full extent. Furthermore, the colour-difference circuits should be reasonably linear, though it is probable that the amounts of non-linearity likely to occur in practice will not give rise to large or intolerable colour errors. It is hoped to confirm this with a subjective experiment at a later date, but it is clear from experience so far that the linearity and tracking requirements of the colour difference circuits will not be so stringent as those determined by Jackson and Jacobs[5] for an RGB system.

3. Circuit Design of Some Decoders

In this Section the circuits of a number of types of colour difference decoders will be described and assessed from the point of view of the circuit designer and the design conditions for operating nominal phosphate and all-sulphide tubes using weighted outputs and for unweighted designs will be given.

Assessment of the colorimetric performance of these decoders from the point of view of errors in component values, etc., will be dealt with in Section 4.

3.1. *The X/Z Decoder*

The so-called X/Z system was introduced by R.C.A. some ten years ago. For derivation of the required colour difference signals from the subcarrier chrominance signal only three colour difference amplifying valves are required over and above the usual pair of synchronous detectors, the matrixing operation being performed by a common cathode resistance. For this circuit to yield the required output signals appropriate demodulation axes (known as X and Z) must be chosen. Unlike the I and Q axes, these X and Z axes do not represent a system specification and are not fixed, but depend upon the matrix circuit parameters and the tube drive requirements. (See Appendix 1.)

A further feature possible with the X/Z decoder is the possible provision of line blanking and d.c. restoration by the addition of an extra triode.

3.1.1. Circuit description

Figure 5 shows the circuit diagram of a typical X/Z decoder. Any type of synchronous detector may

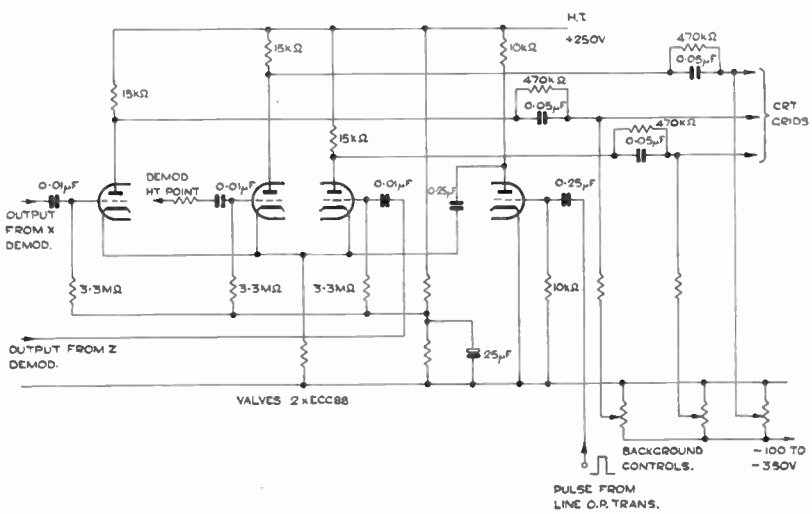


Fig. 5. Typical X/Z decoder circuit.

be used, but triode pulsed-envelope demodulators are usually employed. The three triodes V1, V2 and V3 act as the colour difference amplifiers with cathode coupling. With a low h.t. line of the order of 250 V (which may be preferred on the grounds of circuit cost) the need to supply colour difference output voltages of up to 200 V peak-to-peak makes it necessary to employ low impedance triodes. The maximum anode loads are determined by the required bandwidth and the total driven capacitance (tube grid plus strays). Typically, they may be of the order of 15 kΩ maximum. The triodes used in the circuit built by one of the authors had $R_a = 3{\cdot}5$ kΩ, $g_m = 7{\cdot}7$ mA/V and $\mu = 27$ at the operating point.

The features of the synchronous detectors are not of great importance to the present discussion but it may be noted that the large Miller input capacitance (~ 50 pF) of the red and blue matrix amplifier triodes may be made an integral part of the low-pass filters which follow the X and Z synchronous detectors.

It is shown in Appendix 1 (and also in reference 6) that if the matrix is assumed to have equal anode loads and valve characteristics and the grid of triode V2 is earthed with respect to a.c. signals ($E_2 = 0$) the matrix cannot yield the correct outputs unless the required blue to red tube drive ratio is 0·365 : 1. With any other drive ratio some crosstalk in the outputs and therefore some colour errors must be accepted. Figure 6 shows the errors which occur for a phosphate tube with a gamma of 2 when the crosstalk and amplitude error terms in the outputs are made equal. The demodulation conditions are given below, and the required inputs to the matrix valves V1 and V3 from the X and Z synchronous detectors are found to be in the third quadrant of the N.T.S.C. phasor diagram. It will be seen that the chromaticity errors are small and that the luminance errors are not large.

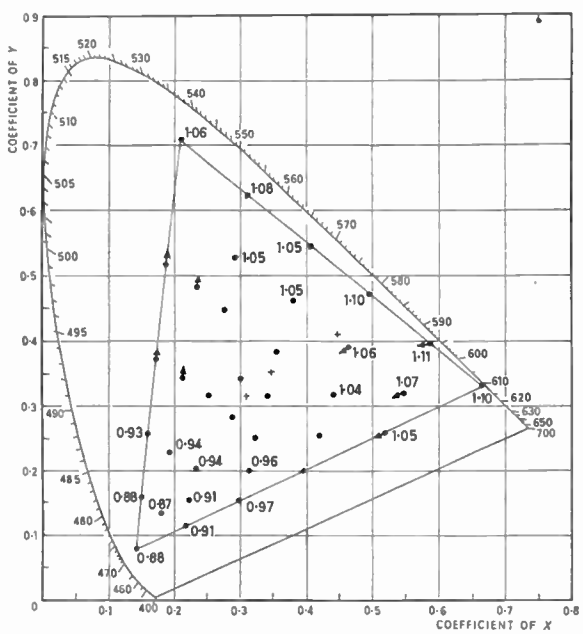


Fig. 6. X/Z decoder errors with $E_2 = 0$ (equal crosstalk and amplitude errors).

It is, however, shown in Appendix 1 that exactly correct matrix circuit outputs can be obtained by making E_2 (the input signal to V2) non-zero and proportional to E_3 (the Z detector output). In this case, of course, the X and Z axes and demodulation gains become different from those when $E_2 = 0$.

In the circuit shown, the d.c. levels are reinserted by means of valve V4. A large amplitude positive line blanking pulse is applied to the grid of V4 from a high impedance source. This causes a negative pulse to be applied to the common cathodes of V1, V2 and V3 which in turn causes each colour difference amplifier to pass grid current during the pulse and the coupling capacitors charge by an appropriate amount. This has the effect of reinserting the d.c. component at the grids. Since the charge acquired, and therefore the grid potential is determined by the total resistance in series with each grid coupling capacitor, steps must be taken to ensure symmetry.

With the d.c. component now reinserted at the grids of the colour difference matrix amplifiers V1, V2 and V3 the anodes can be partially d.c.-coupled to the picture tube grids, and final adjustment of the grid d.c. levels can be made by conventional 'background' controls.

During the line blanking period the grids of the picture tube are driven negative with respect to the cathodes and this automatically provides line flyback suppression.

When a low h.t. line is used it is necessary to return the background controls to a negative supply in order to establish the necessary picture-tube grid-cathode potentials. In a domestic receiver design this may be an undesirable complication.

As an alternative to the technique described above the outputs may be a.c.-coupled to the grids and the d.c. component reinserted at the picture tube grids by means of simple diode clamps driven from the line time-base. As the reinsertion takes place at the last possible point this circuit is less prone to drift than the one using the blanking triode and since, if desired, all the picture tube grids may be clamped to a common potential, setting up of the picture grey-scale tracking is made easier.

In the mathematical analysis of the decoder in Appendix 1 it is shown that the demodulation axes and relative gains, the value of the common cathode

resistor, and the proportion of signal fed to the $(G-Y)$ amplifier grid necessary to give exact colour rendering, depend upon the required picture tube drive ratios, which may differ from tube to tube. Thus, as stated in Section 2.3, it is very desirable to include provision for adjusting the outputs from the colour difference amplifiers. Because of the complex inter-relation of the matrix components this can only be achieved readily by employing potentiometers in the output anode circuits.

From the point of view of simplicity and ease of setting up it would be desirable to design for 'unweighted' outputs at the matrix circuit anodes (cf. Section 2.3) but this can then limit the maximum available output drive to one or more of the picture tube grids. (For example, consider a phosphate display and a decoder operating from a 250 V h.t. line. Then the maximum drive from any amplifier will be, say, about 200 V peak-to-peak. This is the maximum swing possible for the $(B-Y)$ signal, so that on saturated colour bars the maximum drive to the red grid will be $(0{\cdot}7/0{\cdot}89) \times 200 = 156$ V. If, however, the decoder is designed to give the required ratio of phosphate red to blue drive ratio (1 : 0·6) a maximum output swing of 200 V in $(R-Y)$ would necessitate a maximum swing of only $200 \times (0{\cdot}89/0{\cdot}7) \times 0{\cdot}6 = 152$ V peak-to-peak in the $(B-Y)$ output, which is readily obtained. Thus the unweighted design results in a 22% reduction in maximum usable red drive and therefore in brightness.)

3.1.2. Design conditions

Assuming the parameters above, namely N.T.S.C. transmission primaries, Illuminant C white, equal anode loads of 15 kΩ and identical valves with $\mu = 27$, $R_a = 3{\cdot}5$ kΩ, $g_m = 7{\cdot}7$ mA/V, it is possible to calculate the required value of cathode load and the X and Z demodulation phases and gains necessary to drive a tube with a given nominal drive ratio.

As it is possible to obtain exact reproduction by making E_2 non-zero and proportional to E_3 (i.e. to the Z demodulator output) at the cost of two extra components, the case when $E_2 = 0$ will not be considered except to note the required demodulation conditions.

(a) *Design for nominal phosphate tube with* $E_2 = 0$ (5% crosstalk in each output)

The required conditions as derived in Appendix 1 are:

X demodulation phase	= 260·4°
Z demodulation phase	= 197·9°
Relative X/Z demodulation gain	= 1·04
R_K	= 480 ohms

(b) *Exact design for nominal phosphate tube*

The required conditions as derived in Appendix 1 are:

X demodulation phase	= 244·2°
Z demodulation phase	= 197·0°
Relative X/Z demodulation gain	= 0·91
E_2/E_3	= 0·27
R_K	= 612 ohms

(c) *Exact design for nominal sulphide tube with approximate colour point correction*

The required conditions are:

X demodulation phase	= 242·2°
Z demodulation phase	= 195·9°
Relative X/Z demodulation gain	= 0·86
E_2/E_3	= 0·29
R_K	= 611 ohms

(d) *Design for unweighted outputs*

X demodulation phase	= 209·7°
Z demodulation phase	= 193·5°
Relative X/Z demodulation gain	= 0·79
E_2/E_3	= 0·65
R_K	= 1960 ohms

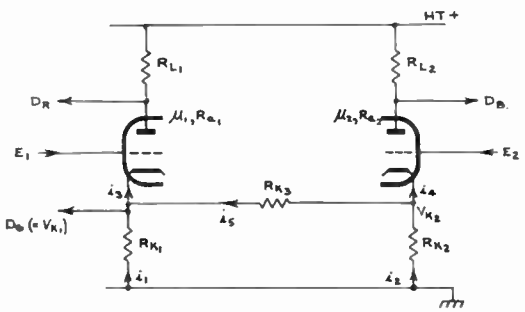
It will be seen that in this last case the X and Z axes are only 16 deg apart, and therefore the demodulation phases and gains will be much more critical than in the previous two designs.

3.2. *The Double-triode Matrix Decoder*

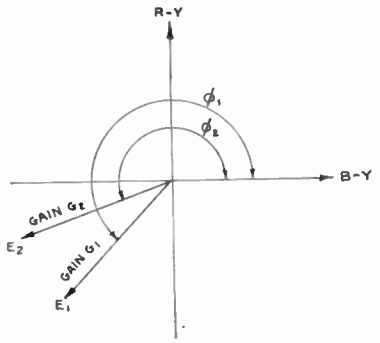
A typical basic design of this type of circuit is shown in Fig. 7(*a*). By accepting some limitation in output drive of the decoder for a given h.t. line it is possible to save a triode function as compared with the X/Z decoder and to derive the required $(G-Y)$ drive signal from the cathode of one of the triodes. It is shown in Appendix 2 that in order to obtain the required colour difference outputs the synchronous demodulators driving the two grids must operate in the third quadrant of the N.T.S.C. phasor diagram (Fig. 7(*b*)) and that some constraint exists between the anode and cathode loads. One disadvantage of this circuit is the fact that if d.c. coupling is employed throughout, the output colour difference signal d.c. levels are not all the same, so that dissimilar brightness control networks are necessary.

As a decoder of this kind has not been constructed by the authors a detailed circuit description will not be given.

However, as an example of the demodulation parameters which arise with this form of decoder



(*a*) Basic double-triode matrix circuit.



(*b*) Demodulator phases and gains.

Fig. 7. Basic circuit of double triode matrix.

matrix they have been determined by means of the theory given in Appendix 2 for a matrix circuit with the following parameters.

$$\mu_1 = \mu_2 = 33$$
$$R_{a1} = R_{a2} = 2{\cdot}6\ \text{k}\Omega$$
$$R_{K1} = 5{\cdot}6\ \text{k}\Omega$$
$$R_{K2} = 4{\cdot}7\ \text{k}\Omega$$
$$R_{K3} = 3{\cdot}3\ \text{k}\Omega$$

(a) *Design for a nominal phosphate tube*

The required conditions are:

$R_{L1} = 8{\cdot}1\ \text{k}\Omega$; $R_{L2} = 7{\cdot}8\ \text{k}\Omega$

Demodulation phase 1	= 239·3°
Demodulation phase 2	= 207·2°
Relative demodulation gain ratio (G_1/G_2) =	0·99

(b) *Design for nominal sulphide tube with approximate colour point correction*

The required conditions are:

$R_{L1} = 8{\cdot}2\ \text{k}\Omega$; $R_{L2} = 8{\cdot}5\ \text{k}\Omega$

Demodulation phase 1	= 239·3°
Demodulation phase 2	= 207·1°
Relative demodulation gain ratio =	0·99

(c) *Design for unweighted outputs*

$R_{L1} = 6{\cdot}5\ \text{k}\Omega$; $R_{L2} = 10{\cdot}4\ \text{k}\Omega$

Demodulation phase 1	= 239·0°
Demodulation phase 2	= 206·7°
Relative demodulation gain ratio =	0·97

3.3. *The (R− Y)/(B− Y) Matrix Decoder*

In this type of decoder the $(R-Y)$ and $(B-Y)$ signals are obtained directly from two synchronous demodulators operating along the relevant axes, which are established 90 deg apart at the transmitter. By taking suitable proportions of the $(R-Y)$ and $(B-Y)$ signals from convenient points in the matrix a positive or a negative $(G-Y)$ signal is obtained and the desired $(G-Y)$ signal is obtained by suitable amplification and/or inversion.

In practice the application of this principle may vary in detail. For example, the $(R-Y)$ and $(B-Y)$ demodulators may operate at high level and drive the picture tube directly, or they may operate at lower level and be used in conjunction with colour difference signal amplifiers. As high level demodulators introduce a number of additional difficulties, they will not be considered in this paper.

For the second approach there are two basic possibilities for the matrixing to obtain the $(G-Y)$ signal. If matrixing of the $(R-Y)$ and $(B-Y)$ signals takes place ahead of the colour difference amplifiers, gain variations in these cannot affect the formation of the $(G-Y)$ signal, but two inverting amplifiers will be necessary. Alternatively, in the interests of economy the matrix may take the high level outputs of the $(R-Y)$ and $(B-Y)$ colour difference amplifiers, but this will involve sacrificing complete immunity to variation in gain of the red and blue colour difference amplifiers.

With regard to the question of demodulator phase accuracy it should be noted that since normal N.T.S.C. subcarrier regenerators produce a reference output which lies in the $(R-Y)$ phase (within the phase accuracy of the regenerator) the $(R-Y)$ signal at least can be obtained with good accuracy since there is no subsidiary phase shift network outside the loop. Moreover, only one phase shift network of 90° is required to establish the other required signal $(B-Y)$.

The circuits to be described may be designed to give weighted or unweighted outputs. Where circuit values are given they are for an unweighted design, but it will be apparent that only minor modifications (of gain and/or matrix resistor values) will be necessary to achieve a particular weighted design. The demodulation axes will remain unchanged which is an attractive feature of this decoder.

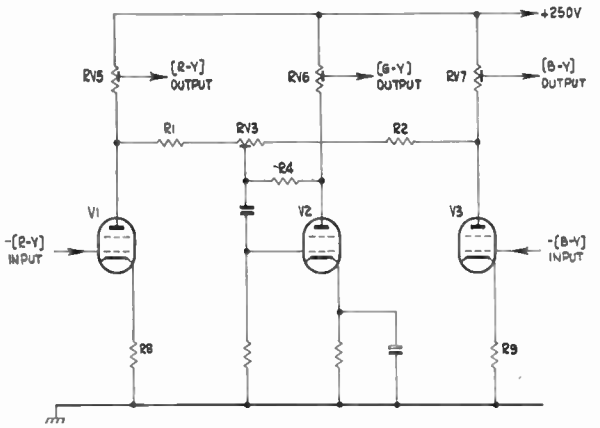


Fig. 8. $(R-Y)/(B-Y)$ decoder with anode matrixing.

3.3.1. Circuits employing matrixing from the anodes of the red and blue colour difference amplifiers

Figure 8 shows the circuit of a combined colour difference amplifier and matrix with the matrixing performed at the anodes of the $(R-Y)$ and $(B-Y)$ amplifiers V1 and V3. These should preferably be pentode valves in order to obtain sufficient output voltage. It is good practice to employ some negative feedback, for example by cathode degeneration, in the interests of gain stability. V2 functions as an anode follower phase inverter and R1, R2 and R4 are chosen to give the correct proportions of $-(E_R'-E_Y')$ and $-(E_B'-E_Y')$ at the output. RV3 is an optional preset control which enables the correct proportions to be obtained even if the $(R-Y)$ and $(B-Y)$ amplifier gains are not quite correct. The large amount of feedback applied round V2 ensures a high degree of gain stability and the virtual earth at the grid effectively prevents crosstalk of the $(R-Y)$ signal into the $(B-Y)$ channel, and vice versa, through R1 and R2.

If a triode valve is used for V2 it may be necessary to compensate for the anode to grid capacitance (which appears in shunt with R4) by shunting R1 and R2 with small capacitors.

The load resistors RV5, RV6 and RV7 are normally about 10–15 kΩ. However, whereas the effective output impedance of V1 and V3 taking into account R1 and R2 may be in excess of 100 kΩ, the output impedance of V2, because of the large amount of feedback, may only be a few hundred ohms. Moreover, any ripple voltages at the anodes of V1 and V3 will be inverted by V2. Thus, this type of circuit can give rise to coloured hum bars in the picture if the h.t. line is inadequately smoothed.

As the theoretical analysis of this type of circuit is very straightforward details will not be given.

3.3.2. A circuit with matrixing before the colour difference amplifiers

A disadvantage of the previous circuit is that variations in the gains of the $(R-Y)$ and $(B-Y)$ amplifiers will in principle influence the $(G-Y)$ signal. Furthermore, drive adjustments for individual tubes have to be made at the output and this may not only upset any frequency compensation but may also result in some restriction of the output drive voltages. By matrixing before the colour difference amplifiers and inserting the drive controls in the difference amplifier grids these difficulties may be avoided at the expense of a further inverting stage.

A typical circuit based on these principles is shown in Fig. 9. The colour difference amplifiers V1b, V2b and V3b have some 12 dB negative feedback to improve stability and reduce the effect of valve spreads. Cathode compensation is used to give a bandwidth of some 2 Mc/s. Individual gain controls with a range of adjustment from zero to 100% to facilitate tracking with the luminance drive controls are incorporated in the grid circuits and since no power dissipation is involved the potentiometers can be quite small.

For the sake of completeness the demodulators and final chrominance amplifier are shown. The complete symmetry of the circuit facilitates neutralizing of the anode-grid capacitances of the chroma amplifier V5 and demodulators V4. Neutralization of the grid-cathode capacitances of V1a and V3a is also possible. Since the final chroma amplifier operates in push-pull to facilitate manual or automatic d.c. control of saturation a quadfilar transformer is necessary to drive the demodulators. This may consist of $4\times 1\frac{3}{4}$ turns of wire on a ferrite bead (Mullard type FX 2249). Once again the analysis of this circuit is quite straightforward.

3.3.3. 'Two pentode' circuits

A further form of circuit employing short suppressor grid base pentodes is shown in Fig. 10.

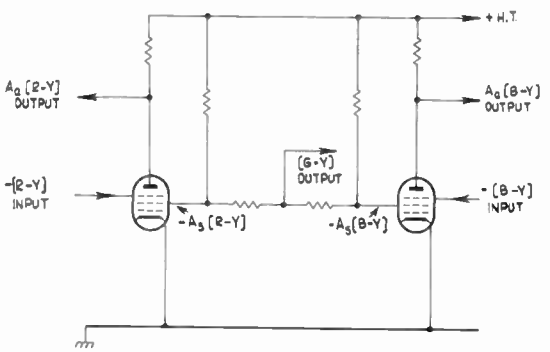


Fig. 10. Principles of colour difference amplifier matrix using two pentodes.

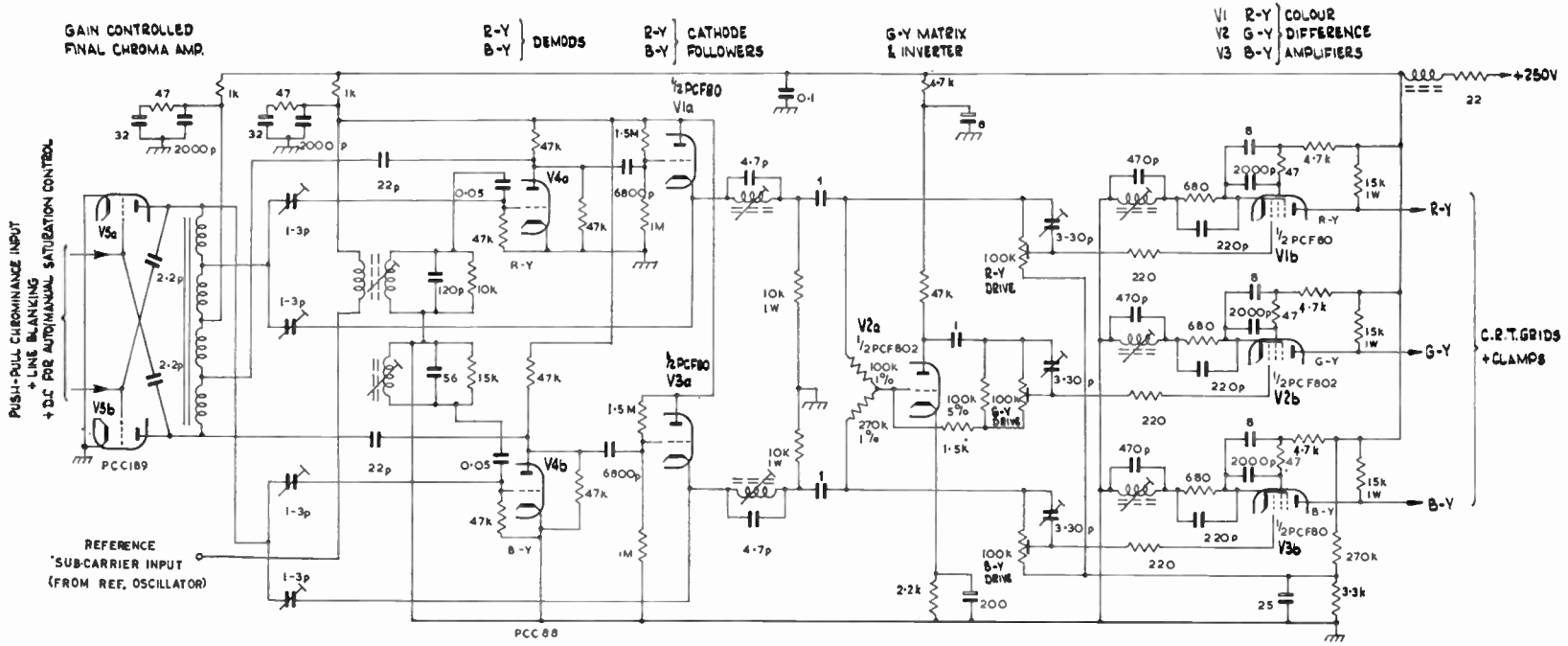


Fig. 9. $(R - Y)/(B - Y)$ decoder with matrixing before the colour difference amplifiers.

All tuned circuits resonant at 4·43 Mc/s.

If $-(R-Y)$ and $-(B-Y)$ signals are fed to the suppressor grids then in-phase amplified signals appear on the screen grids and these may be directly combined to form the $(G-Y)$ signal. (The control grids are arranged to be effectively earthed in this configuration.) It is also possible that the valves could be made to function as demodulators by applying composite chrominance signal to the suppressor grids and reference signal to the control grids. A number of alternative arrangements of this kind may be envisaged. Switched beam-tubes giving outputs of either polarity at the pair of anodes may also be employed.[7]

3.3.4. Decoder with cathode matrixing

The difficulties associated with the $(R-Y)/(B-Y)$ decoders described above, e.g. coloured hum-bars, etc., may be avoided by using a matrix circuit which employs matrixing in the cathode circuits of the colour difference amplifiers. By its very nature such a circuit will give rise to crosstalk in the outputs and ideally for complete elimination of the crosstalk a correcting 'T' network should be connected in the anode circuit. In practice it may be shown that the colour errors may be made negligibly small by inserting a suitable value of resistor common to the anode circuits of the red and blue colour difference amplifiers (see Appendix 3). The circuit thus takes the form shown in Fig. 11.

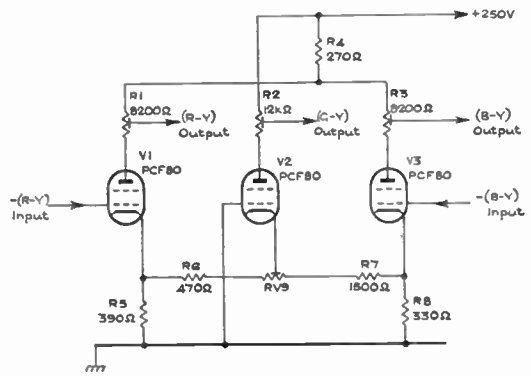


Fig. 11. $(R-Y)/(B-Y)$ decoder with cathode matrixing.

For clarity the screen grid feed arrangements have not been shown, but in a suitable circuit arrangement each valve would have a 4·7 kΩ dropping resistor, which helps to improve gain stability, and a separate 8 μF decoupling capacitor between the screen grid and *cathode*. Similarly the biasing arrangements have not been shown as they will depend upon the choice of valves and the output requirements.

If desired the connection of the green difference signal amplifier cathode to the matrix may be by means of a potentiometer which permits relative adjustment of the contributions of the $(R-Y)$ and $(B-Y)$ signals if one or both of these has incorrect amplitude.

In practice, such an adjustment would appear to be unnecessary. The theoretical analysis of this form of decoder matrix is given in Appendix 3.

Design conditions for unweighted outputs:

Using the pentode portions of PCF80 valves with $g_m = 5{\cdot}5$ mA/V the following design conditions were found by calculation to give unweighted outputs having only negligible crosstalk errors.

R1 = 8·2 kΩ
R2 = 12·47 kΩ (= 12 kΩ + 470 Ω)
R3 = 8·76 kΩ (= 8·2 kΩ + 560 Ω)
R4 = 270 Ω
R5 = 390 Ω
R6 = 470 Ω
R7 = 1·5 kΩ
R8 = 330 Ω

4. Effect of Operational and Component Errors

By means of computer programs it has been possible to determine the colour errors (i.e. errors in chromaticity and luminance) which arise due to operational errors, such as incorrect demodulation phase and gain, and due to errors in component values, which can arise either due to manufacturing spreads or to ageing effects. For simplicity the calculations have been made for a phosphate shadow-mask display assumed to require nominal ratios of red, green and blue drive. (The effects of errors in these were considered in Section 2.3.) However, the results in general will apply substantially to sulphide tubes.

It is further assumed that the transmission and display gammas are 0·5 and 2 respectively and that the N.T.S.C. transmission primaries and Illuminant C white point are employed. Although in practice the display gamma is higher, which will modify the colour errors which occur, this was considered to be a fair basis for the purposes of comparison.

Furthermore, as indicated in Section 2.3, the subjective significance of colour error diagrams of the type to be described has yet to be established, but the relative magnitude of such errors will yield very useful information. To avoid reproducing a large number of colour error diagrams these will only be given for the cases where appreciable errors occur.

4.1. *Errors with the X/Z Decoder*

The errors have been determined for an X/Z decoder with weighted outputs (i.e. in the ratio 1 : 0·8 : 0·6)

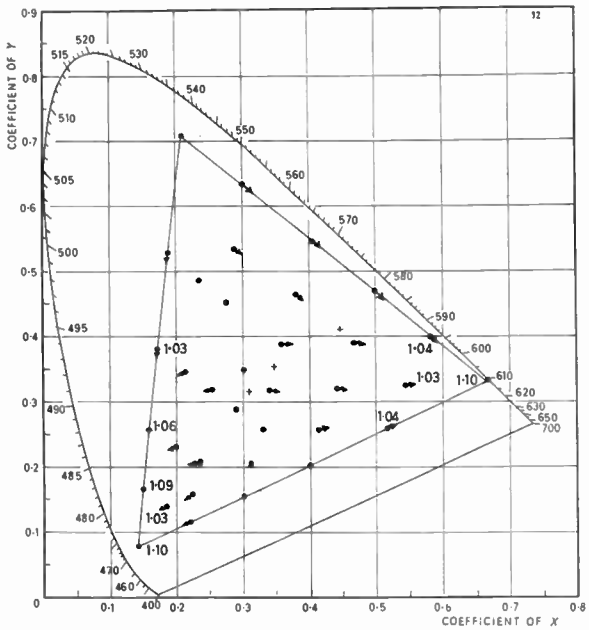


Fig. 12. Colour errors with X/Z decoder due to 5° error in X demodulation phase. (249·2° instead of 244·2°.)

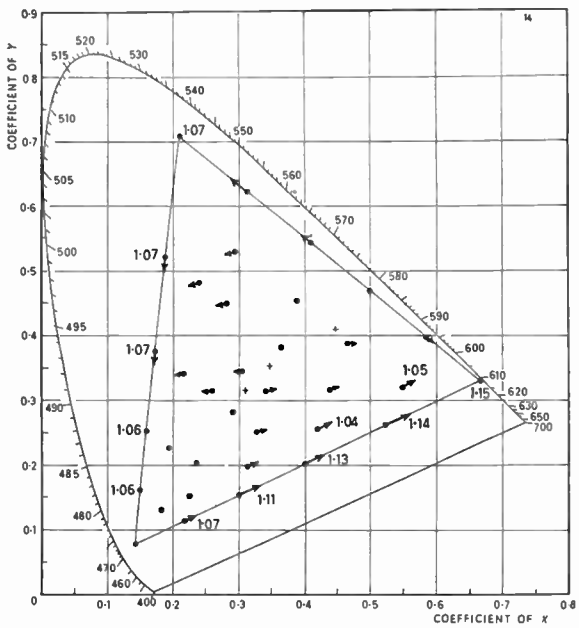


Fig. 14. Colour errors with X/Z decoder due to +10% error in X demodulation gain.

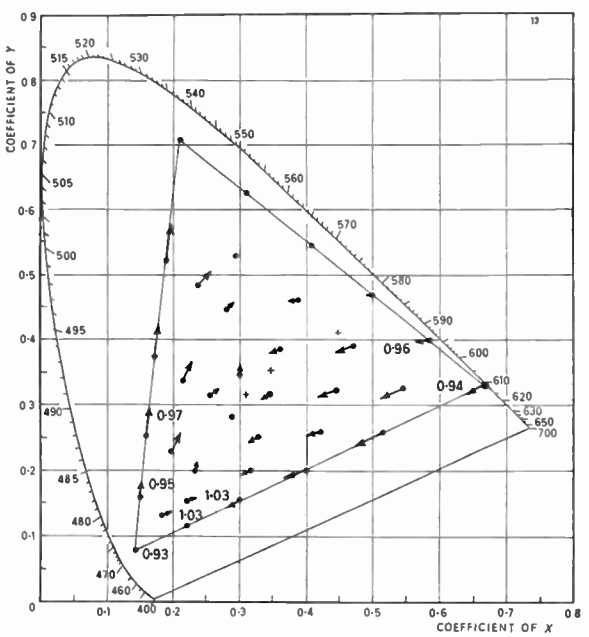


Fig. 13. Colour errors with X/Z decoder due to 5° error in Z demodulation phase. (202° instead of 197°.)

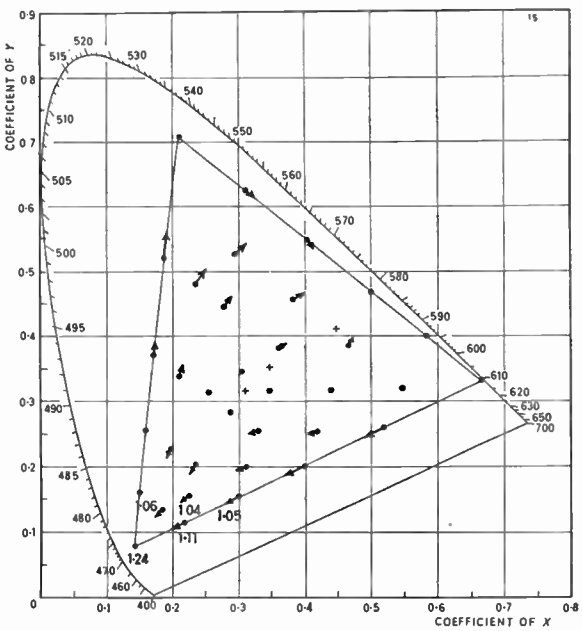


Fig. 15. Colour errors with X/Z decoder due to +10% error in Z demodulation gain.

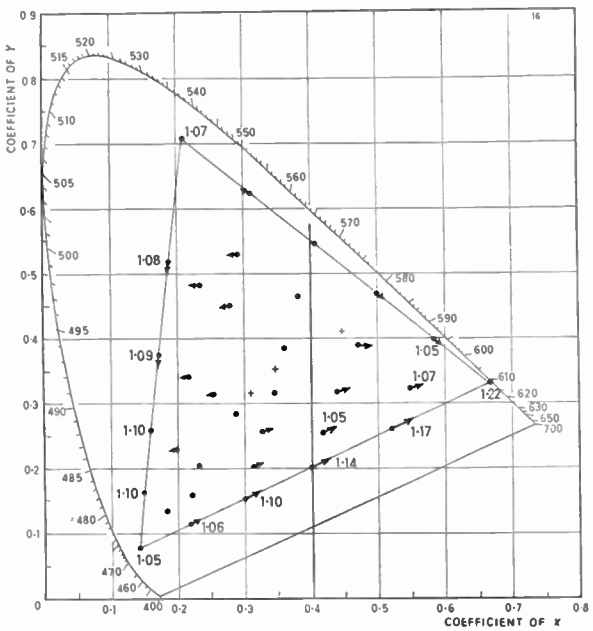


Fig. 16. Colour errors with X/Z decoder due to $+20\%$ error in red triode μ.

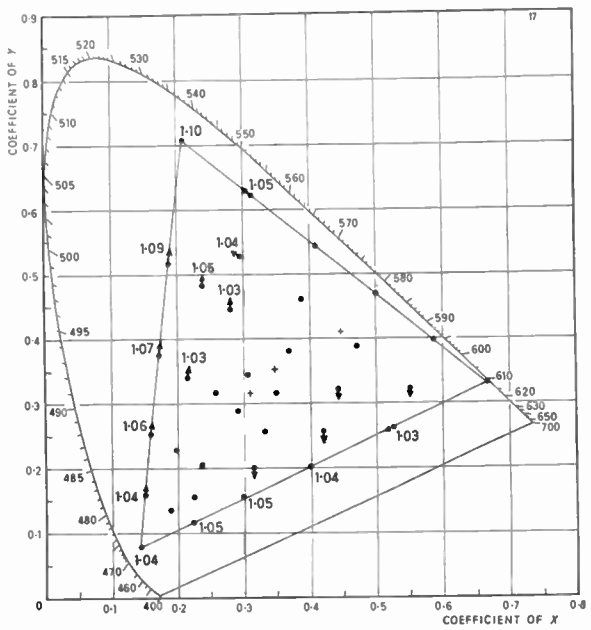


Fig. 17. Colour errors with X/Z decoder due to $+20\%$ error in green triode μ.

according to the design of Section 3.1.2 (b)).

$$\mu = 27{\cdot}0, \qquad R_a = 3{\cdot}5\ \text{k}\Omega, \qquad R_L = 15\ \text{k}\Omega$$

$$\phi_X = 244{\cdot}2^\circ, \qquad \phi_Z = 197{\cdot}0^\circ$$

$$G_{X/Z} = 0{\cdot}91, \qquad E_2/E_3 = 0{\cdot}27, \qquad R_K = 612\ \Omega$$

The effects of $+5^\circ$ errors in the X and Z demodulation phases are shown in Figs. 12 and 13. The errors are not large but a tolerance of less than 5 deg on the Z demodulation phase may need to be set.

The effect of $+10\%$ gain errors in the demodulation gains are shown in Figs. 14 and 15. These are not too large. The effect of a $+10\%$ error in the proportion of the Z demodulator output fed to the green matrix triode is negligibly small.

Figures 16, 17 and 18 show the effects of $+20\%$ errors in μ of the triodes. In particular the appreciable luminance error of blue in Fig. 18 should be noted. As in practice spreads in the values of μ of greater than 20% are not uncommon it will be apparent that these spreads can be a source of large colour errors. As may be expected with triodes with an R_a of 3500 kΩ and anode loads of 15 kΩ the effects of errors in the former are negligible. It may be also shown that the colour errors that can arise due to the use of 10% anode and cathode resistors are insignificant.

It may be noted in general that the situation is much the same for the unweighted design. However, because the demodulation axes are only some 16 deg apart the demodulation gains and phases become even more critical.

4.2. *Errors with the Double-triode Matrix Decoder*

The colour errors occurring due to errors in the unweighted design of double-triode matrix decoder of

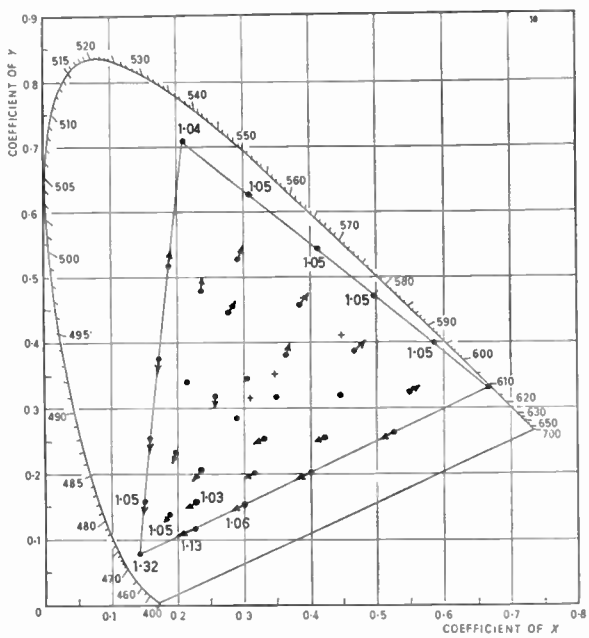


Fig. 18. Colour errors with X/Z decoder due to $+20\%$ error in blue triode μ.

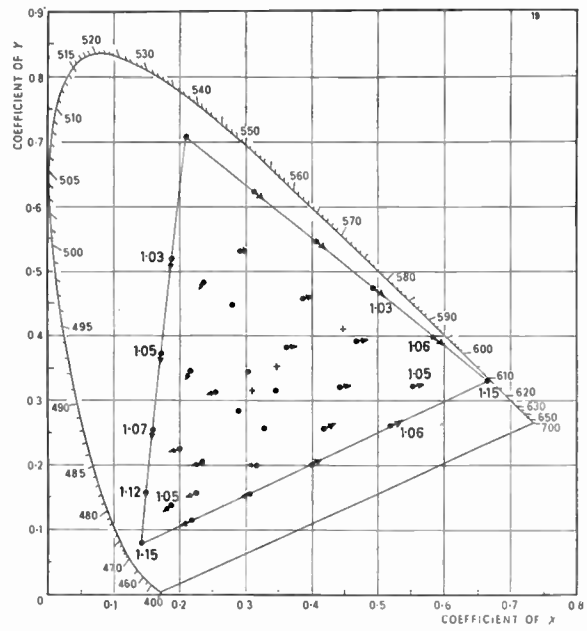


Fig. 19. Colour errors with double-triode decoder due to 5° error in demodulation phase ϕ_1. (244·3° instead of 239·3°.)

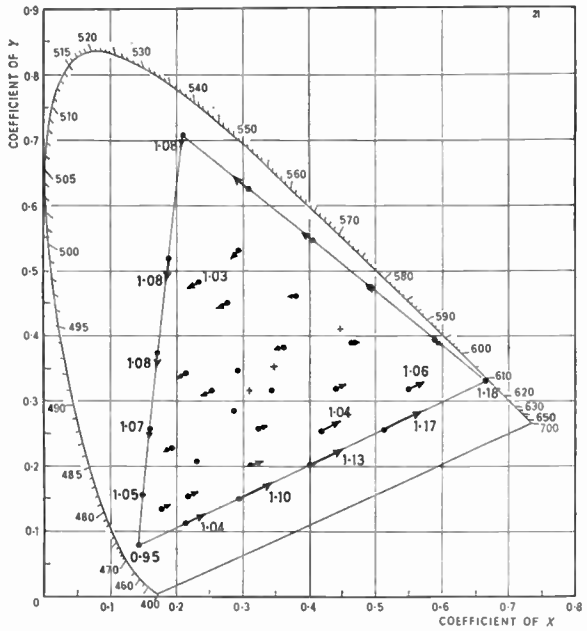


Fig. 21. Colour errors with double-triode decoder due to +10% error in ϕ_1 demodulation gain.

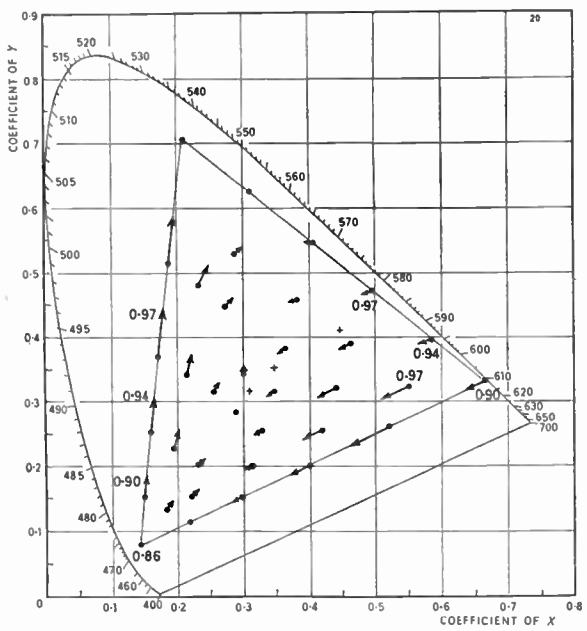


Fig. 20. Colour errors with double-triode decoder due to 5° error in demodulation phase ϕ_2. (212·2° instead of 207·2°.)

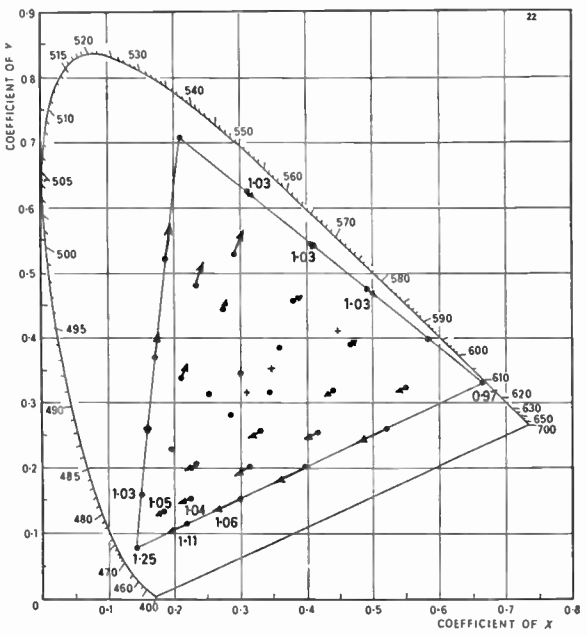


Fig. 22. Colour errors with double-triode decoder due to +10% error in ϕ_2 demodulation gain.

Fig. 7(*a*) for a phosphate display (Section 3.2 (a)) have been determined.

Figures 19, 20, 21 and 22 respectively show the effect of +5 deg errors in the demodulation phases and +10% errors in the demodulation gains. From this it is seen that the phase of the second demodulator (ϕ_2) and the demodulation gains are rather critical.

For 10% errors in the anode or cathode load resistors or 20% errors in triode μ the chromaticity errors are negligibly small and the luminance errors are never very large. Undoubtedly, the insensitivity to errors in μ is due to the fact that the cathode resistors are an order larger in value than the X/Z decoder cathode resistor, and thus produce substantial feedback.

4.3. *Errors with* $(R-Y)/(B-Y)$ *Decoders with Matrixing Before or After the Colour Difference Amplifiers*

The errors in this case were determined for an unweighted design, but the results apply substantially to weighted designs also.

Figures 23 and 24 show the effect of 5 deg errors in the phases of the $(R-Y)$ and $(B-Y)$ demodulators respectively. The errors are seen to be fairly small.

The effect of gain errors in the $(R-Y)$ and $(B-Y)$ channel gains (including the demodulators) have been determined for the two cases of matrixing before and after the colour difference amplifiers. In the former case of course, errors in the $(R-Y)$ or $(B-Y)$ difference amplifier gains will not affect the $(G-Y)$ signal. The errors for 10% gain errors for the two sets of cases are shown in Fig. 25 (*a–d*), whence the chromaticity errors are seen to be small and the luminance errors are not large. The choice of only a 10% error as a basis for calculation was made on the assumption that with this type of decoder it is easy to insert gain adjustments to eliminate most of the effects of initial circuit spreads, which is not the case with the X/Z and double-triode matrix decoders.

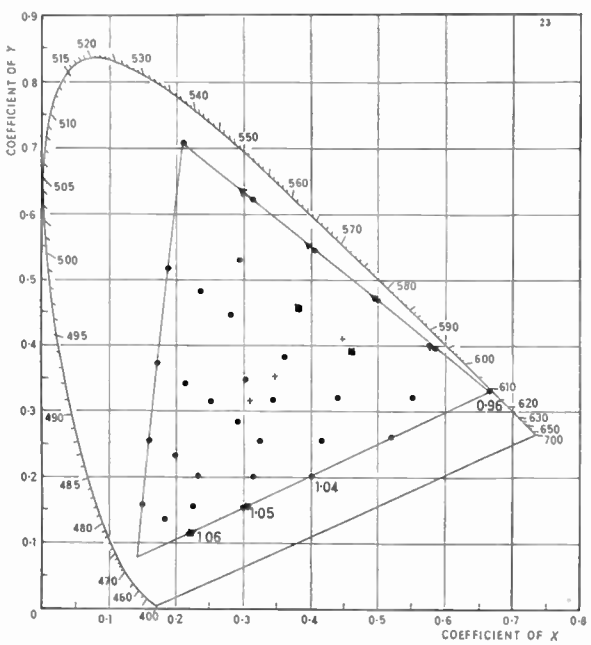


Fig. 23. Colour errors with $(R-Y)/(B-Y)$ decoder due to 5° error in $(R-Y)$ demodulation phase. (85° instead of 90°.)

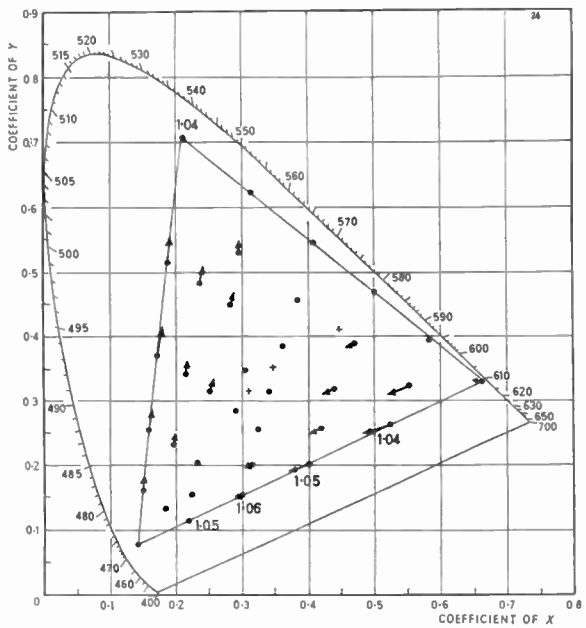
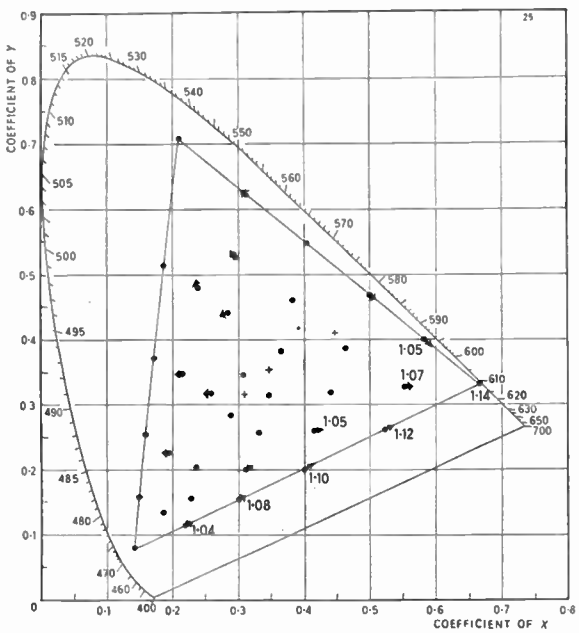


Fig. 24. Colour errors with $(R-Y)/(B-Y)$ decoder due to 5° error in $(B-Y)$ demodulation phase. (5° instead of 0°.)

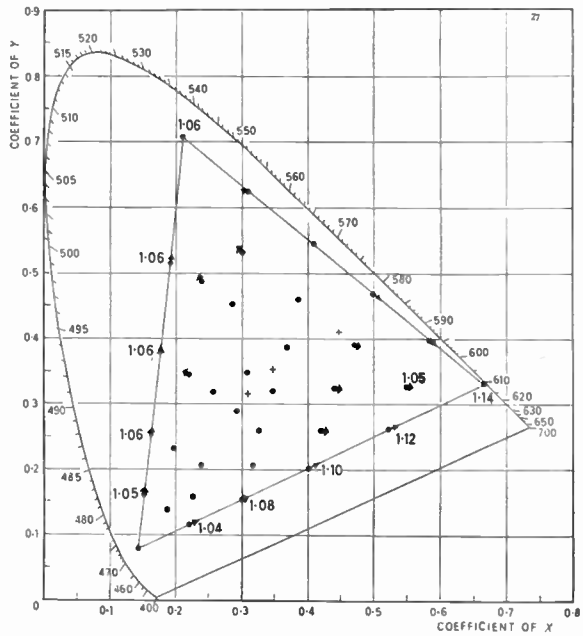
4.4. *Errors with an* $(R-Y)/(B-Y)$ *Decoder with Cathode Matrixing*

Again an unweighted design was assumed with the parameters as given in Section 3.3.4.

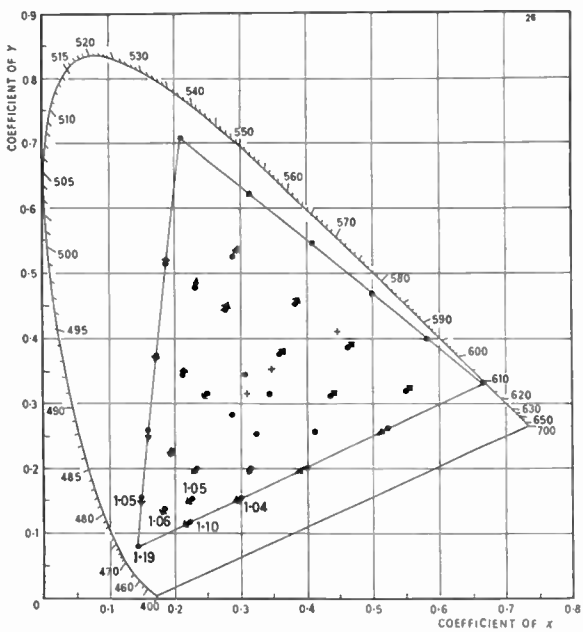
The effects of demodulation phase errors are the same as for the previous case, where they were seen to be small. The effects of demodulation gain errors will be as for the previous case (with the contribution to the $(G-Y)$ signal affected as shown in Figs. 25 (*c*) and (*d*). The effects of 10% errors in the anode and cathode load resistors have been determined and it has been found that the chromaticity errors are small and the luminance errors rarely more than 10%. This is also the case when there is a 20% error in the mutual conductance of one of the difference amplifiers.



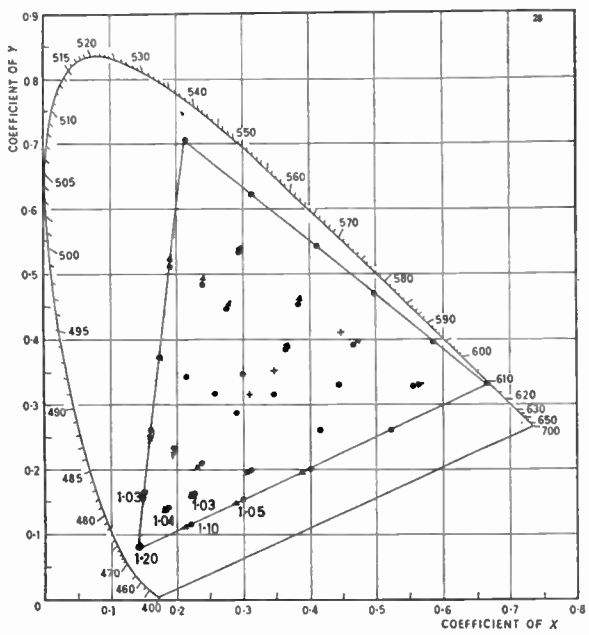
(*a*) 10% error in $(R - Y)$ gain ($G - Y$ signal unaffected).



(*c*) 10% error in $R - Y$ gain (and also in contribution to $(G - Y)$ signal).



(*b*) 10% error in $(B - Y)$ gain ($G - Y$ signal unaffected).



(*d*) 10% error in $B - Y$ gain (and also in contribution to $(G - Y)$ signal).

Fig. 25. Colour errors with $(R - Y)/(B - Y)$ decoder due to errors in $(R - Y)$ and $(B - Y)$ demodulation gains for decoders with matrixing before or after the colour difference amplifiers.

5. Conclusions

From the results of investigations carried out on the X/Z, double-triode and $(R-Y)/(B-Y)$ types of decoders the following deductions may be made.

The X/Z decoder is more critical in respect of the synchronous demodulation phases than the $(R-Y)/(B-Y)$ type of decoder and is also rather critical in respect of the triode gains. Moreover, the complex interdependence of the outputs makes difficult any initial adjustment (including the necessary provision for adjusting the colour difference signal output drives to suit a particular tube). Where a low h.t. voltage is desired the use of a fourth triode for d.c. level reinsertion and blanking is not convenient and it may be difficult to obtain adequate output so that this type of decoder then loses its attractiveness.

For the double-triode matrix type of decoder the matrix configuration need not be very critical, but because of the relatively small angle between the demodulation axes ($\sim 32°$) the demodulation phases and gains are even more critical and must be carefully maintained. Initial adjustment may be difficult and provision must also be made for adjustment of the output drive amplitudes, but the authors have no experience of this problem for this type of decoder.

A further difficulty is likely to arise because the colour difference signals are derived from both anode and cathode circuits. The d.c. output levels may be inconvenient and with a low h.t. line the output drive amplitude will probably be inadequate.

In contrast to the above decoders it has been shown possible to design an $(R-Y)/(B-Y)$ type of decoder which is not critical in respect of any of its component values and is easily designed to permit adjustment of the outputs to suit a particular tube or to compensate for the effect of initial circuit spreads. The cathode matrix type of circuit in particular is a very satisfactory and economical design.

The colour errors occurring in the various decoders are likely in practice to be somewhat different from those calculated here, largely due to the display tube gamma being greater than 2, but the calculations are still considered to give valuable insight into the performance of decoders. Preliminary results of subjective experiments indicate that many of the colour errors of the magnitude described in this paper may in fact be tolerable. However, in practice several parameters may be in error at once and therefore it still seems good design policy to choose the form of decoding which is least critical in respect of component values etc., and can be easily adjusted. In this respect the $(R-Y)/(B-Y)$ type of decoder is most satisfactory and need not involve much additional circuit complexity compared with, say, the X/Z decoder. It is therefore considered to be the most suitable for a domestic receiver.

6. Acknowledgments

Acknowledgment is due to P. S. Carnt and G. B. Townsend for their analysis of the X/Z decoder[6] which served as a useful starting point for the more general analysis given here.

The authors would also like to thank the Director of Mullard Research Laboratories and the Directors of Mullard Ltd. for permission to publish this paper.

7. References

1. P. S. Carnt and G. B. Townsend, "Colour Television—N.T.S.C. Principles and Practice", Chapter 9 (Iliffe, London, 1961).
2. Hazeltine Laboratory Staff, "Principles of Colour Television", Chapter 15 (John Wiley, New York, 1956).
3. R. N. Jackson, "Some factors affecting the choice of a colour television system". Television Society Lecture, April 4th 1963. (To be published in *J. Televis. Soc.*)
4. R. N. Jackson, "Colour corrected decoding for sulphide shadow-mask tubes with three systems of colour television". To be published.
5. R. N. Jackson and T. Jacobs, "An investigation into the subjective effects of some differences between the red, green and blue transfer characteristics of a colour television system", *Acta Electronica*, **2**, Nos. 1–2, p. 95, 1957–8.
6. P. S. Carnt and G. B. Townsend, *op. cit.*, Section 9.12, p. 248 and following.
7. R. Adler and C. Heuer, "Colour decoder simplifications based on a beam deflector tube", *Trans. Inst. Radio Engrs (Broadcast and Television Receivers)*, No. PGBTR-5, p. 64, January 1954.

8. Appendix 1: Theoretical Analysis of the X/Z Decoder

8.1. *General Theory*

The basic circuit, on which the calculations are based is shown in Fig. 26, in which the required outputs are D_R, D_G and D_B. In the normal circuit configuration used the inputs E_1 and E_3 are the

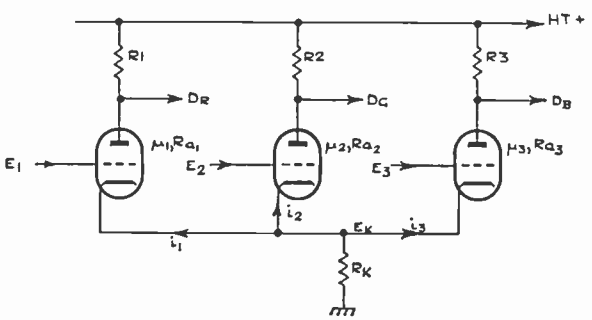


Fig. 26. Basic X/Z decoder matrix circuit.

outputs from the X and Z synchronous demodulators. The input E_2 may be zero but for the present it will be assumed non-zero to conserve symmetry in the analysis.

If the respective triodes have amplification factors μ_1, μ_2 and μ_3, anode impedances R_{a1}, R_{a2} and R_{a3}, the anode loads R_1, R_2 and R_3 and if the valve a.c. signal currents are i_1, i_2 and i_3, the outputs are given by

$$\left.\begin{aligned} D_R &= -R_1 i_1 = -\mu_1'(E_1 - E_K) - E_K \\ D_G &= -R_2 i_2 = -\mu_2'(E_2 - E_K) - E_K \\ D_B &= -R_3 i_3 = -\mu_3'(E_3 - E_K) - E_K \end{aligned}\right\} \quad \text{.....(1)}$$

where

$$E_K = (i_1 + i_2 + i_3)R_K \quad \text{.....(2)}$$

and

$$\mu_1' = \frac{\mu_1 R_1}{R_1 + R_{a1}} \quad \text{etc.} \quad \text{.....(3)}$$

Solution of these simultaneous equations gives:

$$\begin{aligned} D_R &= E_1\mu_1'[KR_2R_3(\mu_1'+1)-1] + \\ &\quad + E_2\mu_2'KR_1R_3(\mu_1'+1) + E_3\mu_3'KR_1R_2(\mu_1'+1) \\ D_G &= E_1\mu_1'KR_2R_3(\mu_2'+1) + \\ &\quad + E_2\mu_2'[KR_1R_3(\mu_2'+1)-1] + \\ &\quad + E_3\mu_3'KR_1R_2(\mu_2'+1) \\ D_B &= E_1\mu_1'KR_2R_3(\mu_3'+1) + E_2\mu_2'KR_1R_3(\mu_3'+1) + \\ &\quad + E_3\mu_3'[KR_1R_2(\mu_3'+1)-1] \end{aligned} \quad \text{.....(4)}$$

where

$$K = \frac{R_K}{\{R_1R_2R_3 + R_K[R_2R_3(\mu_1'+1) + R_1R_3(\mu_2'+1) + R_1R_2(\mu_3'+1)]\}}$$

Now in practice, when designing the decoder, the required outputs D_R, D_G and D_B are given and it is required to find the necessary inputs E_1, E_2 and E_3 to the matrix. This may be achieved by solution of eqns. (4), e.g. by determinants.

8.2. *Simplified Theory*

Although eqns. (4) are required when investigating the errors in the X/Z decoder, they are too complex for use in designing the matrix and it is usual to make the following simplifying assumptions

$$\left.\begin{aligned} R_1 = R_2 = R_3 = R_L \quad \text{(say)} \\ R_{a1} = R_{a2} = R_{a3}, \qquad \mu_1 = \mu_2 = \mu_3 \end{aligned}\right\} \quad \text{.....(5)}$$

so that $\mu_1' = \mu_2' = \mu_3' = \mu'$ (say)

Then eqns. (4) become:

$$\left.\begin{aligned} D_R &= -\mu'[E_1 - L(E_1+E_2+E_3)] \\ D_G &= -\mu'[E_2 - L(E_1+E_2+E_3)] \\ D_B &= -\mu'[E_3 - L(E_1+E_2+E_3)] \end{aligned}\right\} \quad \text{.....(6)}$$

where

$$L = \frac{R_K(\mu'+1)}{R_L + 3R_K(\mu'+1)} \quad \text{.....(6a)}$$

These equations may now be solved to give the E's as follows:

$$\left.\begin{aligned} E_1 &= \frac{-(1-2L)D_R - L(D_G+D_B)}{\mu'(1-3L)} \quad \text{(a)} \\ E_2 &= \frac{-(1-2L)D_G - L(D_R+D_B)}{\mu'(1-3L)} \quad \text{(b)} \\ E_3 &= \frac{-(1-2L)D_B - L(D_R+D_G)}{\mu'(1-3L)} \quad \text{(c)} \end{aligned}\right\} \quad (7)$$

These are the design equations which enable the required inputs E_1, E_2 and E_3 to be determined given D_R, D_G and D_B.

8.3. *Case when* $E_2 = 0$

In the simpler version of the X/Z decoder the input to the second (green) matrix triode is made zero whilst E_1 and E_3 are the direct outputs of the X and Z synchronous detectors.

It will be immediately apparent that in this case the eqns. (7), being three equations in two unknowns, cannot in general be exactly satisfied. The conditions under which they can be satisfied may be determined as follows:

Putting $E_2 = 0$ and subtracting (7b) from (7a) and (7c) yields

$$E_1 = -\frac{1}{\mu'}(D_R - D_G); \qquad E_3 = \frac{1}{\mu'}(D_B - D_G) \quad \text{...(8)}$$

and (7b) may be written as

$$D_G = \frac{-L}{(1-2L)}(D_R + D_B) \quad \text{.....(9)}$$

As a more specific example the case of a phosphate tube will be considered. Then the required outputs are given by

$$\begin{aligned} D_R &= (E_R' - E_Y') \\ D_G &= \alpha(E_G' - E_Y') \\ &= -\alpha[0{\cdot}508(E_R' - E_Y') + 0{\cdot}186(E_B' - E_Y')] \\ D_B &= \beta(E_B' - E_Y') \end{aligned} \quad \text{.....(10)}$$

where 1, α, β are the relative tube drive voltages required to produce the standard white (Illuminant 'C'). Substituting these expressions in (9):

$$D_G = \frac{-L}{(1-2L)}[(E_R' - E_Y') + \beta(E_B' - E_Y')] \quad \text{...(11)}$$

For this equation and the second equation of (10) to be consistent the same relative proportions of $(E_R' - E_Y')$ and $(E_B' - E_Y')$ must occur in each. This is only possible if

$$\beta = \frac{0{\cdot}186}{0{\cdot}508} = 0{\cdot}365$$

Thus with $E_2 = 0$ correct decoding is only possible if the ratio of blue to red gun drives for white is 0·365. In practice, this value is not obtained—in an all-phosphate shadowmask tube, for example, the nominal value of β is 0·6.

In practice, therefore, it is necessary to accept at least one incorrect output signal from the decoder matrix and in general it would appear that the best compromise is to accept small amplitude and cross-talk errors in all three outputs.

Assuming this is done the output signals may be written

$$D_R = (1+x)(E'_R - E'_Y) - l(E'_B - E'_Y) \quad \text{.....(12a)}$$

$$D_G = \alpha[(E'_G - E'_Y) - m(E'_B - E'_Y)] \quad \text{.....(12b)}$$

$$D_B = \beta[(1-y)(E'_B - E'_Y) + n(E'_R - E'_Y)] \quad \text{...(12c)}$$

where x and y are the amplitude errors and l, m, n are the crosstalk errors. No amplitude error is assumed in the $(E'_G - E'_Y)$ component of D_G since this may be eliminated by an overall change of gain. Substituting for $(E'_G - E'_Y)$ in (12b) gives

$$D_G = \alpha[0{\cdot}508(E'_R - E'_Y) + (0{\cdot}186 + m)(E'_B - E'_Y)] \quad (13)$$

But inserting (12a) and (12c) in eqn. (9) gives

$$D_G = \frac{-L}{(1-2L)}[(1+x+\beta n)(E'_R - E'_Y) + \\ + \beta\{(1-y) - l\}(E'_B - E'_Y)] \quad \text{.....(14)}$$

Hence for simultaneous satisfaction of (13) and (14)

$$\frac{1+x+\beta n}{\{\beta(1-y)-l\}} = \frac{0{\cdot}508}{0{\cdot}186+m} \quad \text{.....(15)}$$

Mathematically the optimum solution of this equation, giving minimum errors, occurs when $l = m = n = x = y$. However, from a subjective point of view the colour errors ought to be weighted in terms of their relative visibility. This is much too complex to be considered here, particularly as the design of errorless decoders is possible with E_2 non-zero but, as an example, the case of equal errors will be examined.

Then taking a nominal phosphate tube with $\beta = 0{\cdot}6$, eqn. (15) becomes:

$$\frac{1+1{\cdot}6x}{0{\cdot}6-1{\cdot}6x} = \frac{0{\cdot}508}{0{\cdot}186+x} \quad \text{.....(16)}$$

which has the solution $x \simeq 0{\cdot}05$.

Hence the output signals are

$$D_R = 1{\cdot}05(E'_R - E'_Y) - 0{\cdot}05(E'_B - E'_Y)$$

$$D_G = 0{\cdot}8[(E'_G - E'_Y) - 0{\cdot}05(E'_B - E'_Y)]$$

$$D_B = 0{\cdot}6[0{\cdot}95(E'_B - E'_Y) + 0{\cdot}05(E'_R - E'_Y)] \quad \text{...(17)}$$

These expressions were used to calculate the colour errors shown in Fig. 6.

To determine the demodulation conditions given in Section 3.1.2(a) eqns. (17) may be substituted in (18) to give

$$X \text{ output} = E_1 = \frac{-1}{\mu'}[1{\cdot}456(E'_R - E'_Y) + 0{\cdot}139(E'_B - E'_Y)]$$

$$Z \text{ output} = E_3 = \frac{-1}{\mu'}[0{\cdot}436(E'_R - E'_Y) + 0{\cdot}759(E'_B - E'_Y)]$$

.....(18)

Thus, both demodulation axes lie in the third quadrant of the chrominance phasor diagram. Now over the double sideband region the N.T.S.C. chrominance signal is given by

$$E'_C = \frac{1}{1{\cdot}14}\left[(E'_R - E'_Y)\cos\omega_s t + \frac{1}{1{\cdot}78}(E'_B - E'_Y)\sin\omega_s t\right] \quad \text{.....(19)}$$

Hence the required X demodulation phase is given by

$$\phi_X = \tan^{-1}\left(\frac{1{\cdot}456}{0{\cdot}139} \times \frac{1}{1{\cdot}78}\right) = 260{\cdot}4°$$

and the required Z demodulation phase by

$$\phi_Z = \tan^{-1}\left(\frac{0{\cdot}436}{0{\cdot}759} \times \frac{1}{1{\cdot}78}\right) = 197{\cdot}9°$$

The relative X/Z demodulation gain ratio is given by

$$G_{X/Z} = \frac{1{\cdot}456}{\sin 260{\cdot}4} \Big/ \frac{0{\cdot}436}{\sin 197{\cdot}9} = 1{\cdot}04$$

Given the triode characteristics and the anode loads, the remaining parameter to be determined is the value of the cathode resistor. This may be found by calculating L using, for example, eqn. (14) and then solving eqn. (6a).

8.4. *Case when* $E_2 \propto E_3$

The case when E_2 is made, not zero, but suitably proportional to E_3 will now be considered.

As before the analysis will be carried out for a phosphate display, but it may readily be applied to the sulphide display with approximate colour correction according to the equations of Section 2.4.

For the phosphate tube, substitution for D_R, D_G, D_B in eqns. (7) gives

$$E_1 = \frac{1}{\mu'(1-3L)}[(1-2L-0{\cdot}508\alpha L)(E'_R - E'_Y) + \\ + (\beta - 0{\cdot}186\alpha)L(E'_B - E'_Y)]$$

$$E_2 = \frac{-1}{\mu'(1-3L)}[(L - 0{\cdot}508\alpha + 1{\cdot}016\alpha L)(E'_R - E'_Y) \times \\ \times (L\beta - 0{\cdot}186\alpha + 0{\cdot}372\alpha L)(E'_B - E'_Y)]$$

$$E_3 = \frac{-1}{\mu'(1-3L)}[L(1-0{\cdot}508\alpha)(E'_R-E'_Y)+ +(\beta-2L\beta-0{\cdot}186\alpha L)(E'_B-E'_Y)] \quad \ldots\ldots(20)$$

This gives the extra degree of freedom necessary to enable the correct output colour difference signals to be obtained, and in fact some restriction in the nature of E_2 is possible. A value of L may be chosen such that E_2 is a signal which is readily derived in the circuit. As an example E_2 will be chosen proportional to E_3 in which case

$$\frac{L(1-0{\cdot}508\beta)}{(L-0{\cdot}508\alpha+1{\cdot}016\alpha L)} = \frac{(\beta-2L\beta-0{\cdot}186\alpha L)}{(L\beta-0{\cdot}186\alpha+0{\cdot}372\alpha L)} \quad \ldots\ldots(21)$$

Taking a nominal phosphate tube with $\alpha = 0{\cdot}8$, $\beta = 0{\cdot}6$ the solution of (21) is $L = 0{\cdot}246$ so that eqns. (20) become

$$E_1 = \frac{-1}{0{\cdot}262\mu'}[0{\cdot}402(E'_R-E'_Y)+0{\cdot}109(E'_B-E'_Y)]$$

$$E_3 = \frac{-1}{0{\cdot}262\mu'}[0{\cdot}147(E'_R-E'_Y)+0{\cdot}267(E'_B-E'_Y)]$$

$$E_2 = \frac{-1}{0{\cdot}262\mu'}[0{\cdot}040(E'_R-E'_Y)+0{\cdot}072(E'_B-E'_Y)]$$

$$= 0{\cdot}27E_3 \quad \ldots\ldots(22)$$

giving demodulation conditions

$$\phi_X = 244{\cdot}2^\circ$$
$$\phi_Z = 197{\cdot}0^\circ$$
$$G_{X/Z} = 0{\cdot}91$$

and the value of cathode resistor may be determined by the method outlined above.

9. Appendix 2: Theoretical Analysis of the Double-triode Matrix Decoder

The basic form of circuit used for the calculations on the double triode decoder is shown in Fig. 7.

If it is supposed that V1 and V2 have amplification factors μ_1 and μ_2 and anode impedances R_{a1} and R_{a2} respectively the effective gains are given by

$$\mu'_1 = \frac{\mu_1 R_{L1}}{R_{L1}+R_{a1}}$$

$$\mu'_2 = \frac{\mu_2 R_{L2}}{R_{L2}+R_{a2}} \quad \ldots\ldots(23)$$

where R_{L1} and R_{L2} are the respective anode loads. If the remaining component values and signal components of current and voltage are as shown in Fig. 7 the following relationships must hold.

$$D_R = -\mu'_1(E_1-V_{K1}) = -R_{L1}i_3 \quad \ldots\ldots(24)$$
$$D_B = -\mu'_2(E_2-V_{K2}) = -R_{L2}i_4 \quad \ldots\ldots(25)$$
$$D_G = V_{K1} = R_{K1}i_1 = R_{K3}i_5+R_{K2}i_2 \quad \ldots(26)$$
$$V_{K2} = R_{K2}i_2 \quad \ldots\ldots(27)$$
$$i_3 = i_1+i_5 \quad \ldots\ldots(28)$$
$$i_4 = i_2-i_5 \quad \ldots\ldots(29)$$
$$i_1+i_2 = i_3+i_4 \quad \ldots\ldots(30)$$

Solution of these equations yields

$$E_1 = D_G - D_R/\mu'_1 \quad \ldots\ldots(31)$$

and

$$E_2 = \frac{\mu'_2 R_{L2}R_{K2}D_G-(R_{L2}R_{K2}+R_{L2}R_{K2}+\mu'_2 R_{K2}R_{K3})D_B}{\mu'_2 R_{L2}(R_{K2}+R_{K3})} \quad \ldots\ldots(32)$$

Thus, given the component values and the required values of D_R, D_G and D_B, the required outputs E_1 and E_2 from the synchronous detectors may be determined and hence using the method of Appendix 1, the demodulation gains and phases may be found. For the reverse calculation of the outputs D_R, D_G, D_B to determine the effect of decoder errors given E_1 and E_2 a third equation is necessary.

Since $i_1+i_2 = i_3+i_4$ we have, after multiplying by $R_{K1}R_{L1}R_{L2}(R_{K2}+R_{K3})$:

$$D_R[R_{K1}R_{L2}(R_{K2}+R_{K3})]+ +D_G[R_{L1}R_{L2}(R_{K1}+R_{K2}+R_{K3})]+ +D_B[R_{K1}R_{K2}R_{L1}] = 0 \quad \ldots(33)$$

Hence, given E_1 and E_2, D_R, D_G and D_B may be determined by simultaneous solution of the eqns. (31), (32) and (33).

It will be evident that eqn. (33) must place some constraint between R_{L1}, R_{L2}, R_{K1}, R_{K2} and R_{K3} when D_R, D_G and D_B are given. This constraint will now be established.

In general the required outputs may be written

$$D_R = A_1(E'_R-E'_Y)+A_2(E'_B-E'_Y)$$
$$D_G = A_3(E'_R-E'_Y)+A_4(E'_B-E'_Y)$$
$$D_B = A_5(E'_R-E'_Y)+A_6(E'_B-E'_Y) \quad \ldots\ldots(34)$$

These may be substituted in (33) which must then hold for all values of $(E'_R-E'_Y)$ and $(E'_B-E'_Y)$.

Hence, since R_{L2} cannot be zero, two separate conditions may be found, namely:

$$K_1R_{L2}+K_2R_{L1}R_{L2}+K_3R_{L1} = 0$$
$$K_4R_{L2}+K_5R_{L1}R_{L2}+K_6R_{L1} = 0 \quad \ldots\ldots(35)$$

where

$$K_1 = A_1R_{K1}(R_{K2}+R_{K3}) \quad K_4 = A_2R_{K1}(R_{K2}+R_{K3})$$
$$K_2 = A_3(R_{K1}+R_{K2}+R_{K3}) \quad K_5 = A_4(R_{K1}+R_{K2}+R_{K3})$$
$$K_3 = A_5R_{K1}R_{K2} \quad K_6 = A_6R_{K1}R_{K2} \quad \ldots\ldots(36)$$

Thus, if values are allotted to R_{K1}, R_{K2} and R_{K3} the values of anode loads necessary are fixed in accordance

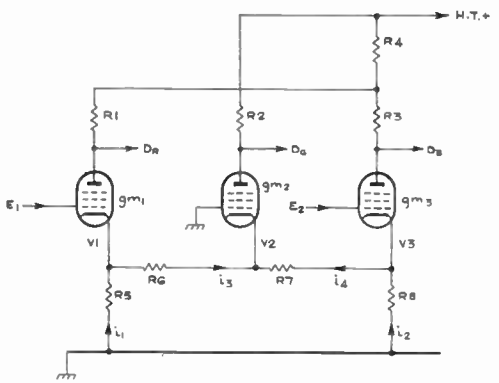


Fig. 27. Basic circuit of $(R-Y)/(B-Y)$ decoder with cathode matrixing.

with the equations:

$$R_{L1} = \frac{K_6K_1 - K_4K_3}{K_3K_5 - K_2K_6}$$

$$R_{L2} = \frac{K_6K_1 - K_4K_3}{K_4K_2 - K_5K_1} \qquad \text{.....(37)}$$

These values may then be inserted in eqns. (31) and (32) to determine E_1 and E_2.

10. Appendix 3: Theoretical Analysis of the $(R-Y)/(B-Y)$ Decoder with Cathode Matrixing

The basic circuit is shown in Fig. 27, and the following equations must be satisfied.

$$D_R = -(R_1+R_4)g_{m1}(E_1-V_1) - R_4g_{m3}(E_2-V_3) \qquad \text{.....(38)}$$

$$D_G = +g_{m2}V_2R_2 \qquad \text{.....(39)}$$

$$D_B = -(R_3+R_4)g_{m3}(E_2-V_3) - R_4g_{m1}(E_1-V_1) \qquad \text{.....(40)}$$

$$V_2 = i_1R_5 + i_3R_6 \qquad \text{.....(41)}$$

$$= i_2R_8 + i_4R_7 \qquad \text{.....(42)}$$

$$V_1 = R_5i_1 \qquad \text{.....(43)}$$

$$V_3 = R_8i_2 \qquad \text{.....(44)}$$

$$i_1 - i_3 = g_{m1}(E_1-V_1) \qquad \text{.....(45)}$$

$$i_2 - i_4 = g_{m3}(E_2-V_3) \qquad \text{.....(46)}$$

$$i_3 + i_4 = g_{m2}V_2 \qquad \text{.....(47)}$$

$$i_1 = V_1/R_5 \qquad \text{.....(48)}$$

$$i_2 = V_3/R_8 \qquad \text{.....(49)}$$

Thus, if D_R, D_G and D_B are given and g_{m1}, g_{m2} and g_{m3} are assumed known, there are 12 equations in the 9 unknowns V_1, V_2, V_3, i_1 to i_4, E_1, E_2 and therefore the resistor values cannot be independent, i.e. there must be some constraint between the values of the resistances.

In practice it is convenient and desirable to choose the demodulation phases to be those corresponding to $(E'_R-E'_Y)$ and $(E'_B-E'_Y)$, and in a particular decoder design it was also convenient to make

$$E_1 = -\alpha(E'_R-E'_Y)$$

and

$$E_2 = -\alpha(E'_B-E'_Y) \qquad \text{.....(50)}$$

where α is a constant.

In theory it is possible to determine the constraints between the resistance values set by the above equations. For example, given R_5, R_6, R_7 and R_8 it is possible to determine R_1, R_2, R_3 and R_4. In practice, a satisfactory solution can rapidly be arrived at by means of the computer program used to investigate the effect of component errors.

Solution of the preceding equations yields the following relationships:

$$\frac{V_1}{R_5} + \frac{V_1-V_2}{R_6} = g_{m1}(E_1-V_1) \qquad \text{.....(51)}$$

$$\frac{V_3}{R_8} + \frac{V_3-V_2}{R_7} = g_{m3}(E_2-V_3) \qquad \text{.....(52)}$$

$$\frac{V_2-V_1}{R_6} + \frac{V_2-V_3}{R_7} = -g_{m2}V_2 \qquad \text{.....(53)}$$

Given R_5, R_6, R_7, R_8, g_{m1}, g_{m2}, g_{m3} these may be solved to give V_1, V_2 and V_3. Then substituting for V_1, V_2 and V_3 in (38), (39) and (40) gives equations for D_R, D_G and D_B in terms of E_1, E_2 and the g_m's and R's.

Proceeding in this way for given values of g_{m1}, g_{m2}, g_{m3} and of R_5, R_6, R_7 and R_8 it was then possible using the computer program to determine the resulting colour errors for particular values of R_1, R_2, R_3 and R_4.

By this means it was rapidly found that for an unweighted design of the decoder (giving equal drive outputs) and with $g_{m1} = g_{m2} = g_{m3} = 5$ mA/V and with $R_5 = 390\,\Omega$, $R_6 = 470\,\Omega$, $R_7 = 1{\cdot}5$ kΩ and $R_8 = 330\,\Omega$ the following values of resistors gave errors in chromaticity too small to be shown on the C.I.E. diagram and errors in luminance within the accuracy of the computer program (i.e. $< 2\%$).

$$R_1 = 8{\cdot}2\ \text{k}\Omega$$

$$R_2 = 12{\cdot}47\ \text{k}\Omega = 12\ \text{k}\Omega + 470\ \Omega$$

$$R_3 = 8{\cdot}76\ \text{k}\Omega = 8{\cdot}2\ \text{k}\Omega + 560\ \Omega$$

$$R_4 = 270\ \Omega$$

Manuscript first received by the Institution on 27th February 1964 and in final form on 4th May 1964. (Paper No. 916/T27.)

Processing Data from N.A.S.A. Satellites

New equipment has now been installed at the D.S.I.R. Radio Research Station, Slough to process the raw data received from the second Anglo-American satellite (UK-2/S-52) launched by N.A.S.A. on 27th March, 1964, and from the third satellite in this series. The equipment performs the initial processing of the telemetry tapes before the data are passed to computer centres elsewhere for further processing, sorting and mathematical analysis.

At Winkfield (Berkshire), one of the 12 radio installations located throughout the world, the position and direction of satellites launched by N.A.S.A. are recorded on tape. The Winkfield installation also makes regular recordings on magnetic tape of the UK-2/S-52 telemetry signals. Additional telemetered data are recorded at the overseas laboratories of the Radio Research Station at Singapore and Port Stanley (Falkland Islands). All three stations are equipped with 'tele-command' facilities; radio transmitters send signals to the satellite to relay recorded measurements in the satellite back to Earth.

The magnetic-tape records from the N.A.S.A. network of receiving stations and from Singapore and Port Stanley are sent to the Radio Research Station at Slough, where the recorded information is converted into a form suitable for feeding into fast digital computers.

The measurements are recorded in a pulsed frequency-modulation code. The average pulse repetition rate is 55 per second while the signal frequency of each pulse, representing the magnitude of the experimental quantity concerned, falls within the band 5 to 15 kc/s. At the receiving station, the time of reception is recorded continuously in a coded form on two other tracks alongside the data from the satellite.

The tapes received at the Radio Research Station are first 'inspected' in one section of the processing system. The parts which have an adequate signal/noise ratio are selected and their usefulness is assessed.

Selected telemetry tapes are then replayed into the main section of the processing system. This essentially digitizes in binary form the experimental magnitudes represented by the signal frequencies of the pulses. The signals are first separated from the background noise by a bank of 128 filters. When a frequency falls within the pass band (100 c/s wide) of a filter the output from this filter suppresses the noise outputs from all the other filters. Thus signals as much as 12 decibels below noise level can be recovered. Each filter is connected to its own binary number generator and the number, in binary-coded decimal form corresponding to the active filter, is passed to the output. Conversion accuracy from frequency to digital number is about 1%.

A measurement made in the satellite can be continuously variable or in the nature of a count. If it is continuously variable the data are interpreted in binary-coded decimal form as described above. For a measurement where counting is performed in the satellite, this is done in binary form and converted to an octal (0–7) scale before transmission to the ground station. Provision is therefore made in the processing equipment for interpreting these data in binary-coded octal form.

A pulse generator in the form of a digital electronic flywheel adjusts itself to the incoming pulse repetition frequency. Data processing continues to generate sampling pulses at the last known pulse repetition frequency when short-period 'fades' occur in the telemetry signal. The flywheel also operates the decimal-octal counter, which is automatically set to the correct position in the pre-arranged sampling pattern by special synchronizing pulses contained in the telemetry signals.

While the measurements from the satellite are being converted to a digital form by the processing equipment the time information on other tracks of the telemetry tape is being decoded and converted into binary-coded decimal form. This decoded time is held in a register which is continuously updated so that it is accurate to the nearest millisecond. Time readings are taken from this register on the occurrence of synchronizing pulses in the telemetered data.

The experimental measurements and time, both in digital form, are assembled in a magnetic-core buffer store. When the store is full the information is rapidly read out on to magnetic tape in a form suitable for input to a digital computer. An automatic checking procedure is incorporated in the processing equipment so that errors introduced into the data during processing are immediately detected.

The equipment has been designed so that it can also deal with measurements, which have been made during a whole orbit of the earth, stored on a tape recorder inside the satellite and replayed rapidly on command from a ground station. The latter measurements can be automatically selected by the equipment, using special 'start' and 'end' pulses, contained in the telemetry signals. An extra fixed frequency is also present between the pulses of stored data. This is digitized in the same way as the data and enables corrections to be made in the subsequent computer analysis, to compensate for variations in the speed of the satellite tape recorder.

Acknowledgment

This article has been compiled from information, received from D.S.I.R., on the data processing equipment designed and supplied by Plessey (U.K.) Limited.

U.D.C. 621.394.45

A Wideband Spectrum Analyser-Synthesizer for the Prediction and Smoothing of Interrupted Speech

By

P. V. INDIRESAN, Ph.D.
(*Associate Member*)†

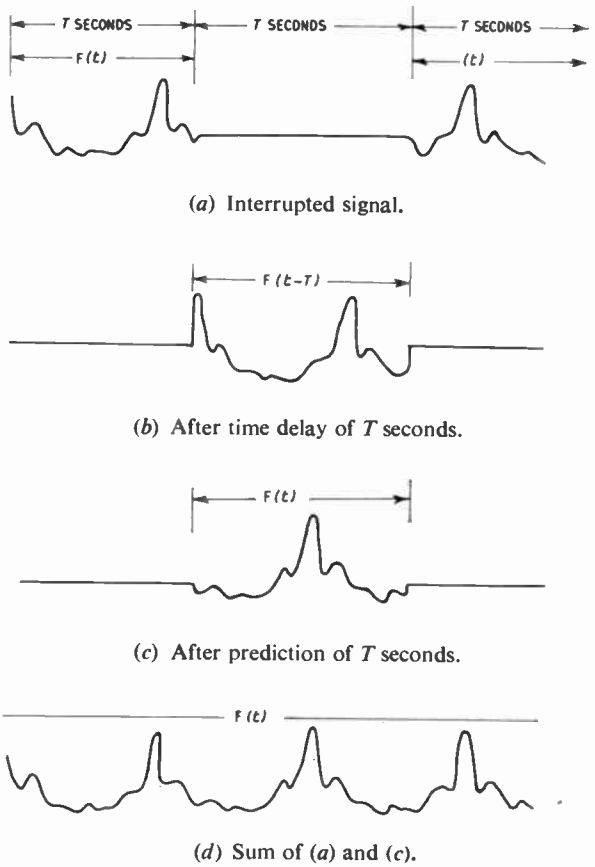
Presented at the Symposium on "Telecommunications and Electronics" sponsored jointly by the Centre of Advanced Study in Radio Physics and Electronics of the University of Calcutta and the Calcutta Zone of the Indian Division of the Institution on 29th February 1964.

Summary: Time sharing of slowly-interrupted speech will be a possible alternative to the vocoder for increasing the capacity of a speech communication channel, if the associated switching noise can be eliminated. It has been shown that this noise can be suppressed by reiterating the interrupted speech samples at the receiver so that they fill in the blank spaces correctly both in time and in phase. A wideband phase advancing network for speech signals is required for such a coherent reiteration and a practical technique has been proposed for this purpose. The error inherent to prediction and the additional error introduced by the practical limitations of the proposed system have been estimated. The system described can help to increase the line capacity by a factor of two or four, or eight at the most. The wideband phase correction network described here is essentially a spectrum analyser-synthesizer and can be used for the analysis and synthesis of any quasi-time-invariant complex wave.

1. Introduction

Speech interrupted at a slow rate of a few tens of cycles per second has been found experimentally to be intelligible.[1] As the signal is off for part of the time additional signals can be interpolated during these silent intervals, and as the switching frequency is low there is no appreciable increase in bandwidth due to switching. Thus, without increasing the overall bandwidth, several signals can be transmitted by time-division multiplexing and thereby the communication rate can be increased. However, interrupted speech cannot be used directly as it suffers from switching noise. Three methods have been proposed[2] for overcoming this noise, (a) by a system of artificial switched reverberation, (b) by integration and (c) by coherent reiteration. Reiteration is the process by which interrupted speech samples are delayed in time and interpolated into the silent intervals in such a manner as to produce a continuous signal. Such reiteration, in addition to filling the blank space correctly in time, should also be in the proper phase or else the sudden change in phase will produce a transient noise and there will be no improvement in spite of having a signal all the time.[3] The process by which the reiteration is exact both in time and in phase is called here coherent reiteration.

A method for obtaining the coherent reiteration of interrupted speech is shown schematically in Fig. 1.



(*a*) Interrupted signal.

(*b*) After time delay of T seconds.

(*c*) After prediction of T seconds.

(*d*) Sum of (*a*) and (*c*).

Fig. 1. Smoothing of interrupted speech by coherent reiteration.

† Telecommunication Engineering Department, University of Roorkee, India.

The original signal is interrupted for exactly half the time so that the sampling and interruption periods are both equal to, say, T seconds. If such interrupted speech is delayed by T seconds, it will fill exactly the blank spaces in the undelayed signals. However, its frequency components will be retarded by ωT radians where ω is the angular frequency of the component in question. This delayed signal is therefore passed through a network which advances the phase of each component by ωT. Such a network is called a predictor. Such a predicted signal will fill in the blank space correctly both in time and in phase. Thereby, the original continuous signal will be reproduced without error, provided it has not changed during the period of prediction. The heart of this system is the predictor. Predictors for simple signals of known characteristics are well established but, no method appears to have been developed so far for complex signals like those of speech. A prediction and phase correction technique suitable for such complex signals is discussed here.

2. A Prediction of Wideband Signals

The future value $F(t+T)$ of an input signal $F(t)$ for a prediction period of T seconds can be expressed by the Fourier integral

$$F(t+T) = 1/2\pi . \int_{-\infty}^{+\infty} e^{j\omega t}(t+T)\,d\omega . \int_{-\infty}^{+\infty} F(\tau).e^{-j\omega\tau}\,d\tau \qquad .....(1)$$

In eqn. (1), the integration over time for values of τ greater than t corresponds to future values of the input function $F(t)$. No physically realizable system can be clairvoyant and take into account events that occur in the future. Hence, the best physically-realizable prediction is given by

$$F_1(t+T) = 1/2\pi . \int_{-\infty}^{+\infty} e^{j\omega(t+T)}\,d\omega . \int_{-\infty}^{t} F(\tau).e^{-j\omega\tau}\,d\tau \qquad .....(2)$$

Consequently, there will be an unavoidable error in prediction and an exact solution is impossible. Nevertheless, due to the inertia which is inevitable to all physical systems, there is a limit to the rapidity with which any change can be made and, what is more, to the rapidity with which the change can take effect. Speech, for example, is the output of the various resonant parts of the vocal system, and due to their narrow bandwidth the sounds produced remain virtually invariant for small intervals of time. This effect is further enhanced by the reverberation in the room in which the sound is heard. For instance, if the reverberation time is 0·6 second, each of the frequency components of speech is effectively passed through a filter with a bandwidth of ± 2 c/s only. Consequently, although speech signals occupy a band of 4000 c/s or more, they can vary only very slowly and each of their frequency components, once produced, will persist as long as a tenth of a second. It is because of this short-time invariancy of speech that interrupted speech is at all intelligible. Hence, although even the best theoretically realizable system does introduce some prediction error, the effect is not appreciable for periods of the order of a few milliseconds. The magnitude of this error has been estimated in Appendix 1 to be given by

$$[E(TN)]^2 = [E(T)]^2/[F(t)]^2 = 1-e^{-2\alpha T} \qquad ...(3)$$

where $[E(T)]^2$ = mean square error for prediction period T,

$[E(TN)]^2$ = normalized mean square error, and

α = decay constant of the voice system consisting of the source, receiver and channel and taking into account reverberation in the room in which the sound is produced and the room in which it is heard.

Even if the rooms are fairly dead (reverberation time of about 0·5 seconds) the value of α will be of the order of 10 nepers per second and the error due to prediction about 10% for a prediction period of 10 milliseconds. This calculation was made on the basis that random sounds are being pronounced. In normal speech, on the average the error will be two to three times less. However, mean-squared error is not of much direct use in determining the quality of speech transmission which can be evaluated only by articulation tests. These figures for mean-squared error are given only to indicate the order of error introduced.

3. A Predictor for Wideband Signals

As the speech signal is a real-valued function, equation (2) may be rewritten as

$$\begin{aligned} F_1(t+T) &= 1/2\pi . \left[\int_{-\infty}^{+\infty} \cos\omega(t+T).d\omega \int_{-\infty}^{t} F(\tau).\cos\omega\tau\,d\tau + \right. \\ &\qquad \left. + \int_{-\infty}^{+\infty} \sin\omega(t+T).d\omega \int_{-\infty}^{t} F(\tau).\sin\omega\tau\,d\tau \right] \\ &= 2\int_{0}^{\infty} \cos\omega(t+T).df \int_{-\infty}^{t} F(\tau)\cos\omega\tau\,d\tau + \\ &\qquad + 2\int_{0}^{\infty} \sin\omega(t+T).df \int_{-\infty}^{t} F(\tau)\sin\omega\tau\,d\tau \qquad ...(4) \end{aligned}$$

where $f = \omega/2\pi$.

A practical predictor should evaluate equation (4) accurately. The exact evaluation is, in fact, not practicable, but the integration over time may be

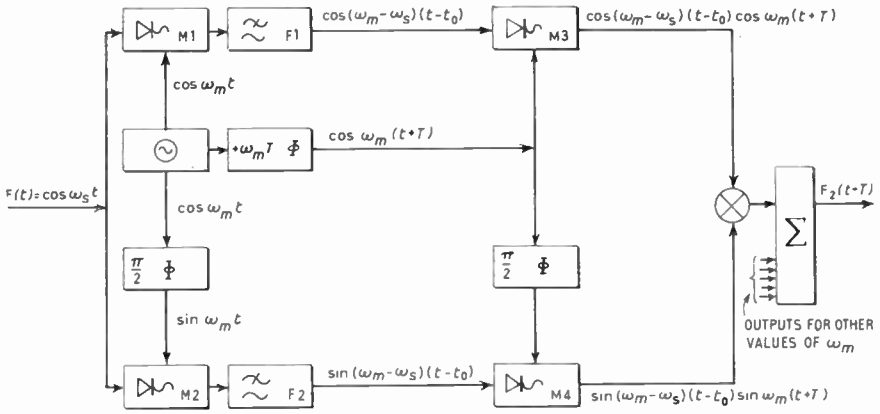


Fig. 2. Schematic diagram of spectrum analyser-synthesizer system for wideband phase correction and prediction.

approximated by a low-pass filter and the integration over frequency approximated by a summation for a set of discrete frequencies spread over the signal frequency band. Such an arrangement has been shown schematically in Fig. 2. For the sake of simplicity, the operation has been indicated for one of the frequencies ω_m used for the summation. If n different frequencies are used for the summation, n networks of the type shown in Fig. 2 will be required. The resultant output will then be of the form

$$F_2(t+T) = \sum_1^n \cos\omega_m(t+T).\int_{-\infty}^{t} F(\tau).\cos\omega_m t.\frac{1}{t_0}\times$$

$$\times e^{-(t-\tau)/t_0}.d\tau + \sin\omega_m(t+T).\int_{-\infty}^{t} F(\tau)\times$$

$$\times \sin\omega_m t.\frac{1}{t_0} e^{-(t-\tau)/t_0}.d\tau \quad ...(5)$$

where t_0 is the time-constant of the low-pass filters F1 and F2.

In Fig. 2 the multiplication of $F(t)$ with $\cos\omega_m(t+T)$ and $\sin\omega_m(t+T)$ is performed by the modulators M1 and M2 respectively, and the resultant product integrated by the low-pass filters F1 and F2. It is interesting to note that the phase advancing is done by shifting the fixed frequency tones ω_m. This needs only a fixed adjustment of a phase shifter. Equation (5) is approximately the same as equation (4) when the time-constant t_0 of the low-pass filters is large. Thus, the scheme of Fig. 2 will produce approximately the prediction value of the input function $F(t)$. The prediction period T can be adjusted to any desired value by choosing the appropriate value of phase advance for the n switching tones. The error introduced by approximating the integration over frequency of equation (4) by the summation of equation (5) has been calculated in Appendix 2. It has been shown that for $F(t) = \cos\omega_s t$,

$$F_2(t+T) = A\cos[\omega_s(t+T)-\theta] \quad ...(6)$$

where

$$A = a_{\Delta\omega_0} + 2\sum a_{k\Delta\omega}\cos k\Delta\omega(t_0+T)$$

$$\theta = \Delta\omega_0(t_0+T)$$

$\Delta\omega$ = the spacing of the frequencies ω_m

$\Delta\omega_0 = \omega_0 - \omega_s$ = the smallest value of $\omega_m - \omega_s$, that is, the difference between ω_s and the switching frequency ω_0 which is closest to it

$a_{\Delta\omega_0} = a_{\omega_0-\omega_s} = \dfrac{1}{\sqrt{1+(\Delta\omega_0 t_0)^2}}$ = the amplitude response of the low-pass filters at the frequency $\Delta\omega_0$, and

$a_{k\Delta\omega} = \dfrac{1}{\sqrt{1+(k\Delta\omega_0 t_0)^2}}$ = the amplitude response of the filters for the kth switching frequency from ω_0.

In equation (6) the amplitude response A is essentially a constant except for one single term

$$a_{\Delta\omega_0} = a_{\omega_0-\omega_s}$$

which only depends on ω_s. A may therefore be taken to be practically a constant. The phase error becomes a maximum for the maximum value of $\Delta\omega_0$ which is $\Delta\omega/2$, that is

$$\theta_{max} = \Delta\omega/2(t_0+T)$$

For a given permissible maximum phase error this sets a limit to the frequency separation of the switching frequencies, and thereby for any desired bandwidth determines the minimum number of frequencies required for the summation. Thus, if B is the band-

width to be covered in radians per second,

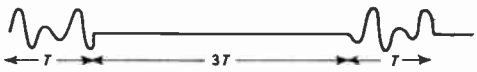
$$n = B/\Delta\omega = B(t_0+T)/2\theta_{max}$$

The phase error θ will cause a reiteration error proportional to $\cos\theta$, and to keep the average error below 1 %, θ_{max} should be limited to 1/5th of a radian. On this basis, to cover a telephone band of 4000 c/s n becomes about a hundred for a prediction period of 10 milliseconds. Then a hundred uniformly spaced ω_m will be needed in the summation and a hundred networks of the type shown in Fig. 2 will be required for prediction. For periods of prediction greater than 10 milliseconds a proportionately larger value of n will have to be used.

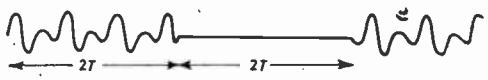
It was assumed in eqn. (5) that the bandwidth of the low-pass filter was small. In practice this implies that the bandwidth is small enough to suppress the sum frequency component $\omega_m+\omega_s$ produced by the modulators M1 and M2. With eqn. (6) it was assumed that the bandwidth is large compared to $\Delta\omega$ so that the time delay remains constant. In practice, with n of the order of a hundred, $\omega_m+\omega_s$ will be much larger than $\Delta\omega$ and both conditions can be satisfied. It can also be verified that a decrease in the bandwidth reduces the time delay at higher values of $\omega_m-\omega_s$ and that θ decreases whereas the amplitude response becomes less uniform. Better performance can be obtained by using more complex filters than the simple R-C network assumed here.

4. Repeated Reiteration

The discussion so far has been for the case in which the sampling and interruption periods were equal. Such interrupted speech can only double the transmission rate. For greater improvement, the ratio of interruption period to sampling period should be correspondingly larger. This ratio cannot be increased by reducing the sampling period as it should be long enough to contain at least one fundamental period of the speech wave, which goes up to 10 milliseconds. Hence the communication rate can be increased only by increasing the interruption period beyond the 10 milliseconds assumed so far. Such an increase will need a proportionately larger number of frequencies for the summation and consequently the complexity and cost of the equipment will increase in direct proportion. An even greater difficulty is the rapid increase in the inherent error of prediction. Both from theoretical considerations[5] and by experiment[1] it has been found that the error becomes too large for interruption periods greater than 100 milliseconds. This sets a limit of ten to the increase in communication rate that one may expect to get. Figure 3 shows a method by which the technique of synchronous reiteration can be extended to increase the communication rate by a factor of four. The reiteration is done in two steps. First, the signal which is on for T seconds and off for $3T$ seconds is reiterated for a period of T seconds to yield a new interrupted signal with equal on and off periods of $2T$ seconds. This new signal is then smoothed to give a continuous signal by reiterating it for a period of $2T$ seconds. In the same manner the technique can be extended to the case where the sample period is one-eighth of the total switching period to give an improvement in communication efficiency of eight times.



(*a*) Input interrupted signal.



(*b*) After coherent reiteration for time T.

(*c*) After a second coherent reiteration for time T.

Fig. 3. Smoothing of interrupted speech whose sample duration is a quarter of the switching period.

5. Conclusion

The prediction technique described here is essentially a means for computing the frequency analytical form of the Fourier integral for a future value of the input signal. This spectrum analysis is performed by using a set of n oscillators whose frequencies are uniformly distributed over the signal frequency range. The use of a two-phase system helps to preserve the phase information in the input signal, and only pre-adjusted fixed phase shifters are needed for the phase-advancing required for prediction. Thus, the need to analyse the instantaneous frequency of the input signal and to advance the phase of such a variable frequency function is avoided. There are two types of error in the proposed system: (a) an unavoidable error caused by any change in the input signal during the period of prediction and (b) a practical error caused by the fact that the integration over frequency is simulated by a summation. The latter determines the number of frequencies required for the summation. It has been estimated that for telephone signals this number will be about a hundred for a prediction period of 10 milliseconds. The former

limits the maximum period permissible for prediction to a tenth of a second, preferably much less. The predictor described here can be used to smooth out the switching noise associated with interrupted speech by the process of synchronous reiteration and thereby make it suitable for communication. By time division several such interrupted speech signals can be sent over a given channel, and as the slow rate of switching does not appreciably increase bandwidth this results in an increase in the communication rate. The use of synchronous reiteration to increase transmission by factors of two, four and eight have been described. Compared to vocoder systems interrupted speech does not alter the formant frequencies in the original signal and hence should give a much more natural reproduction. In addition, interrupted speech systems will have the advantages of time division compared to frequency division that will have to be used with vocoders. It should also be pointed out that the phase adjustment technique described here is very general in its application and can be used for any type of phase correction for complex quasi-time-invariant signals.

6. References

1. G. A. Miller and J. C. R. Licklider, "Intelligibility of interrupted speech", *J. Acoust. Soc. Amer.*, **22**, pp. 167–73, March 1950.
2. P. V. Indiresan, "Interrupted speech and the possibility of increasing communication efficiency", *J. Acoust. Soc. Amer.*, **35**, pp. 405–8, March 1963.
3. D. J. Sharf, "Intelligibility of reiterated speech", *J. Acoust. Soc. Amer.*, **31**, pp. 423–7, April 1959.
4. Y. W. Lee, "Statistical Theory of Communication", Chapter 15 (Wiley, New York, 1960).
5. H. S. Black, "Modulation Theory", p. 50 (Van Nostrand, Princeton, 1953).
6. J. H. Laning and R. H. Battin, "Random Processes in Automatic Control", pp. 225–31 (McGraw-Hill, New York, 1956).

7. Appendix 1: Error Inherent in Prediction

The speech signal $F(t)$ may be expressed in terms of the impulse response $W(t, \tau)$ and the input control function $x(\tau)$ of the vocal source. That is,

$$F(t) = \int_{-\infty}^{t} W(t_1 \tau).x(\tau).d\tau = \int_{0}^{\infty} W(\tau).x(t-\tau).d\tau$$

and

$$F(t+T) = \int_{0}^{\infty} W(\tau).x(t+T-\tau).d\tau$$

$$= \int_{T}^{\infty} W(\tau).x(t+T-\tau).d\tau + \int_{0}^{} W(\tau).x(t+T-\tau).d\tau \quad ...(7)$$

The second term in eqn. (7) refers to future values of $x(t)$ and is the unpredictable part of $F(t+T)$. It therefore is the prediction error $E(T)$. Then the normalized mean-squared error becomes

$$\overline{[E(TN)]}^2 = \frac{\overline{[E(T)]}^2}{\overline{[F(t)]}^2} = \frac{\overline{\left[\int_{0}^{T} W(\tau).x(t+T-\tau).d\tau\right]^2}}{\overline{\left[\int_{0}^{\infty} W(\tau).x(t+T-\tau).d\tau\right]^2}}$$

which becomes[6]

$$\overline{[E(TN)]}^2 = \frac{\int_{0}^{T} W(\tau_1)d\tau_1 \int_{0}^{T} W(\tau_2).\phi_{xx}(\tau_1-\tau_2).d\tau_2}{\int_{0}^{\infty} W(\tau_1)d\tau_1 \int_{0}^{\infty} W(\tau_2).\phi_{xx}(\tau_1-\tau_2).d\tau_2} \quad .....(8)$$

where $\phi_{xx}(\tau_1-\tau_2)$ is the autocorrelation coefficient of $x(t)$.

The exact solution of eqn. (8) is impossible as the characteristics of acoustical systems are too complex to be evaluated precisely. However, a reasonably good approximation is to assume the system to behave like an infinite set of tuned filters all of the same decay constant α nepers per second corresponding to a bandwidth of $\pm\alpha$ radians. Then $W(\tau) = e^{-\alpha\tau}$ and $\phi_{xx}(\tau) = e^{-\alpha\tau}$. Substituting these values eqn. (8) becomes

$$\overline{[E(TN)]}^2 = \frac{\int_{0}^{T} e^{-\alpha\tau_1}d\tau_1\left[\int_{0}^{\tau_1} e^{-\alpha\tau_2}.e^{-\alpha(\tau_1-\tau_2)}.d\tau_2 + \int_{\tau_1}^{T} e^{-\alpha\tau_2}.e^{-\alpha(\tau_1-\tau_2)}.d\tau_2\right]}{\int_{0}^{\infty} e^{-\alpha\tau_1}d\tau_1\left[\int_{0}^{\tau_1} e^{-\alpha\tau_2}.e^{-\alpha(\tau_1-\tau_2)}.d\tau_2 + \int_{\tau_1}^{T} e^{-\alpha\tau_2}.e^{-\alpha(\tau_1-\tau_2)}.d\tau_2\right]}$$

$$= 1 - e^{-2\alpha T} \quad .....(9)$$

8. Appendix 2:
Error Due to Summation with a Finite Set of Frequencies

From Fig. 2, the output $F_2(t+T)$ corresponding to an input $\cos \omega_s t$ is

$$\begin{aligned} F_2(t+T) &= \sum a_{\omega_s-\omega_m} \cos[\omega_s(t+T)-(\omega_s-\omega_m)(t_0+T)] \\ &= a_{\Delta\omega_0} \cos[\omega_s(t+T)-\Delta\omega_0(t_0+T)] + a_{\Delta\omega_0-\Delta\omega} \cos[\omega_s(t+T)-(\Delta\omega_0-\Delta\omega)(t_0+T)] + \\ &+ a_{\Delta\omega_0+\Delta\omega} \cos[\omega_s(t+T)-(\Delta\omega_0+\Delta\omega)(t_0+T)] + a_{\Delta\omega_0-2\Delta\omega} \cos[\omega_s(t+T)-(\Delta\omega_0-2\Delta\omega)(t_0+T)+ \\ &+ a_{\Delta\omega_0+2\Delta\omega} \cos[\omega_s(t+T)-(\Delta\omega_0+\Delta\omega)(t_0+T)] + \ldots \quad \ldots\ldots(10) \end{aligned}$$

If the bandwidth of the filters is large compared to $\Delta\omega$, then

$$a_{\Delta\omega_0-k\Delta\omega} \simeq a_{\Delta\omega_0+k\Delta\omega} \simeq a_{k\Delta\omega}$$

and eqn. (10) reduces to

$$\begin{aligned} F_2(t+T) &\simeq [a_{\Delta\omega_0} + \sum 2a_{k\Delta\omega} \cos k\Delta\omega(t_0+T)] . \cos[\omega_s(t+T)-\Delta\omega_0(t_0+T)] \\ &= A \cos[\omega_s(t+T)-\theta] \quad \ldots\ldots(11) \end{aligned}$$

where $$A = a_{\Delta\omega_0} + \sum 2a_{k\Delta\omega} \cos k\Delta\omega(t_0+T)$$

and $$\theta = \Delta\omega_0(t_0+T) \quad \ldots\ldots(12)$$

Manuscript first received by the Institution on 6th November 1963 and in final form on 1st May 1964. (Paper No. 917.)

U.D.C. 621.375.3

A Direct Analogue of a Magnetic Amplifier

By

R. E. KING, Ph.D., M.Sc.†

Summary: The paper describes the development from first principles of a direct electronic analogue of a series-connected self-excited magnetic amplifier. From a closed loop network representing a simple saturable reactor, the introduction of additional windings and feedback to the basic transductor is developed in analogue form.

The analogue is a simple and accurate aid to the explanation of the principles of saturable reactors and their application in magnetic amplifiers. The ability to observe the effects of varying any of the parameters, e.g. relative number of turns of the various windings, core permeability, self-excitation and bias, is one of the assets of the analogue both for design as well as for educational purposes.

1. Introduction

Analogue computer techniques in analysis, design and synthesis have, over the past years, been proved invaluable. Where mathematical formulation and solution is difficult or impossible it is usually possible to render a reasonable approximation of a complex physical process by simulating it element by element. This, the 'direct analogue method' has enormous advantages for both obtaining a good understanding of the *modus operandi* of the system under investigation as well as being invaluable in determining optimum parameter settings.

The magnetic amplifier is a particular problem which can be conveniently and easily set up on an analogue computer and this paper describes the development of a series-connected self-excited magnetic amplifier. All the quantities associated with a physical magnetic amplifier are directly represented and the analogue has been found invaluable in the study of both steady-state as well as transient performance[1, 2, 3] of this element.

2. Development of the Direct Analogue

2.1. *The Basic Saturable Reactor*

With reference to Fig. 1 in which a simple saturable reactor is placed in series with a resistance R_L, the instantaneous current i_a in the circuit gives rise to an m.m.f., $N_a i_a$, and flux ϕ in the coil. The resultant back e.m.f. e_b in the coil will oppose any change of current and is given by

$$e_b = -N_a \frac{d\phi}{dt} = -N_a A \frac{dB}{dt}$$

where N_a is the number of turns of the coil and A the cross-sectional area of the core. Similarly a voltage $e_L = e - e_b = R_L i_a$ appears across the resistance. Hence the applied e.m.f. is

$$e = R_L i_a + N_a \frac{d\phi}{dt}$$

$$= R_L i_a + N_a^2 \frac{d\phi}{dH} \cdot \frac{di_a}{dt}$$

$$= R_L i_a + L_a \frac{di_a}{dt}$$

where $H = N_a i_a$ the core m.m.f.

and $L_a = N_a^2 \frac{d\phi}{dH}$, the self-inductance of the coil.

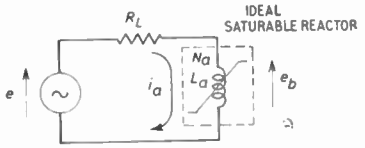


Fig. 1. The saturable reactor.

For a magnetization characteristic which is idealized as being linear up to the saturating level or knee, H_S,

$$\frac{d\phi}{dH} = \frac{\phi}{H} = A\mu_r\mu_0 \qquad \text{for } H \leqslant H_S$$

where $\mu_r\mu_0 = \mu_i$ the incremental permeability

μ_0 = permeability of air

μ_r = relative permeability of core material

and $\frac{d\phi}{dH} = 0$ for $H > H_S$.

For a non-linear characteristic

$$e_b = -N_a \frac{d\phi}{dt} = -N_a A \frac{dB}{dt}$$

† Department of Electrical Engineering, The Queen's University of Belfast.

The direct analogue of this element can thus be produced using conventional high-gain d.c. computing amplifiers and is shown in Fig. 2. The saturation characteristic representing the magnetization curve of the core material has for convenience unity slope between the saturation levels $\pm H_S$. The output signal β of this function is thus directly proportional to the flux density. In the physical system

$$e_b = -N_a A \frac{dB}{dt}$$

and in the analogue

$$e_b = -RC \frac{d\beta}{dt}$$

thus the two are related by

$$RC = N_a A \mu_r \mu_0$$

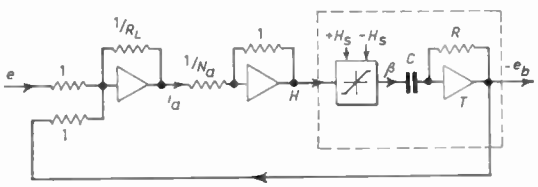


Fig. 2. Saturable reactor analogue.

Further, if L is the self-inductance of the core for small signals, i.e. for linear operation then the time-constant of the differentiator

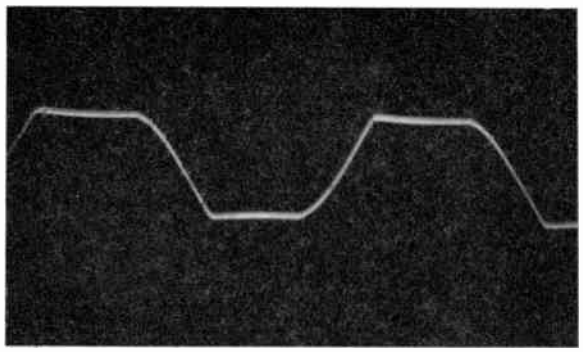
$$T = RC = \frac{L}{N_a}$$

Any instability of the differentiator may be corrected by the addition of a small resistance r in series with C in which case the transfer function of the differentiator is strictly

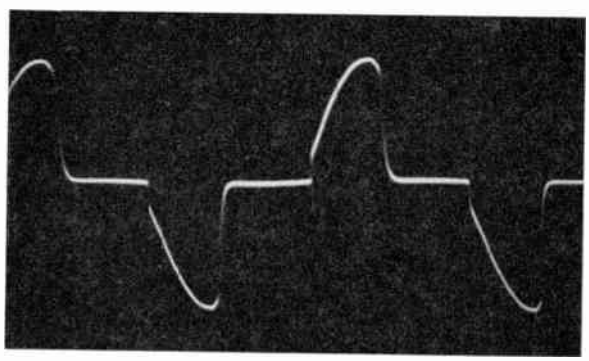
$$D(p) = \frac{RCp}{1+rCp}; \; p = d/dt$$

r can be made suitably small so that the break-point angular frequency $1/rC$ is sufficiently high to pass the highest harmonic required, say $r < 0{\cdot}05R$.

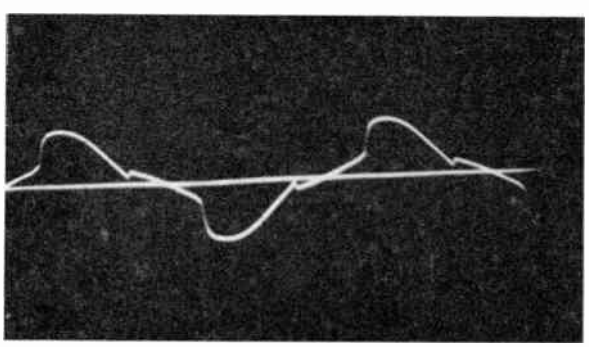
The analogue of Fig. 2 therefore represents the operation of a simple saturable reactor. The effect of variations in any parameter can be simply observed. Changes in the shape of the B–H characteristic can be readily effected by substitution of the appropriate function in place of the simple saturation characteristic. Thus a hysteretic or 'on/off' type can be substituted and a study of waveforms made. Examples of flux density, back e.m.f. and load current for the simple transductor for the idealized saturation type B–H characteristic are shown in Fig. 3.



(*a*) Flux.



(*b*) Voltage across reactor.



(*c*) Load current.

Fig. 3. Waveforms from saturable reactor analogue.

2.2. *The Biased Saturable Reactor (Half-wave)*

The ability to control the mean load current $\bar{i}_a$ by means of a control signal winding provides an introduction to the topic of magnetic amplifiers. The addition of the d.c. signal, or bias, to the basic circuit of Fig. 1, resulting in the formation of a 'transductor', is shown in Fig. 4.

In this case the operating point or mean level of the core m.m.f. can be controlled by the m.m.f. due

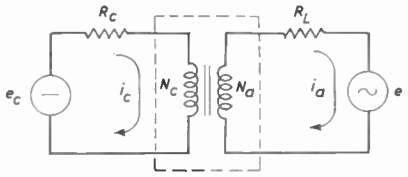


Fig. 4. The basic transductor.

to the mean control current $\bar{i}_c$. Thus to the analogue of Fig. 3 it is simply necessary to add a term $i_c N_c$ to the core m.m.f. H. Due to transformer action from the secondary to the primary an additional term must, however, be included in deriving the control current. Reflected into the control circuit the back-e.m.f. e_b is equivalent to $\left(\frac{N_c}{N_a}\right) e_b$ and

$$R_c i_c = e_c + \left(\frac{N_c}{N_a}\right) e_b$$

The analogue of the basic transductor is now that of Fig. 5.

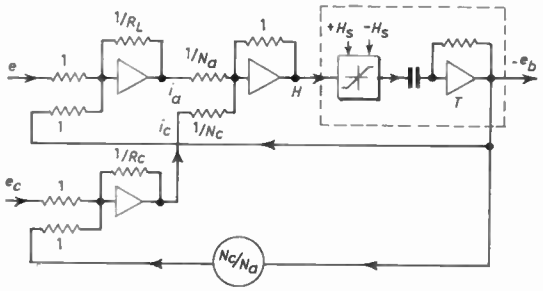


Fig. 5. Analogue of transductor.

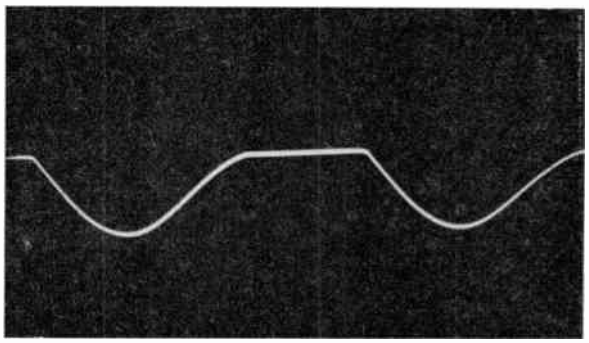
All the relevant waveforms can be observed and changes to any of the transductor parameters be demonstrated. The variation of mean load current $\bar{i}_a$ with changes in the d.c. biasing signal e_c can be made using average-reading instruments directly connected to the outputs of the relevant amplifiers.

In this and all analogues in this paper it is assumed that both the a.c. and d.c. sources have negligible source impedance.

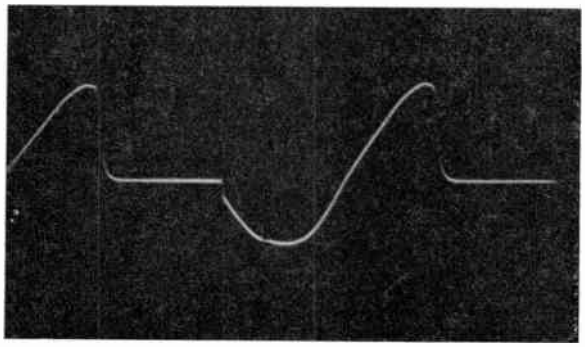
Examples of waveforms obtained are shown in Fig. 6.

2.3. *Full-wave or Balanced Transductor*

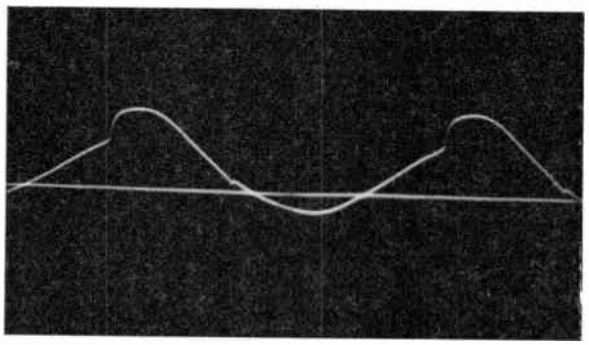
With the basic arrangement discussed in Section 2.2 the transformer action due to the mutually-coupled control and load windings will induce a detrimental a.c. component of fundamental frequency into the control winding thereby disturbing the amplifying properties of the transductor.



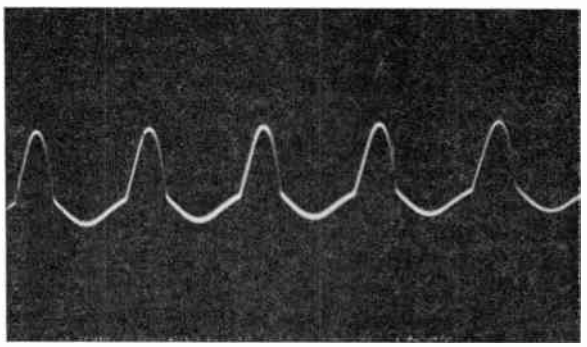
(*a*) Flux.



(*b*) Voltage across reactor.



(*c*) Load current.



(*d*) Total core m.m.f.

Fig. 6. Waveforms from analogue of transductor.

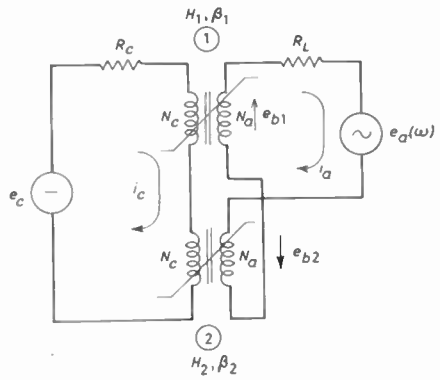


Fig. 7. Full-wave transductor.

The coupling effect can be largely overcome by the use of twin series-opposed windings as shown in Fig. 7. This arrangement provides cancellation of the a.c. component in the control circuit. When a control signal e_c is applied, a current i_c flows through the two cores setting up m.m.f.'s of magnitude i_cN_c in each. The total m.m.f.'s in the two cores will therefore be of the form.

$$H_1 = i_cN_c \pm i_aN_a \qquad \text{for core 1}$$
$$H_2 = i_cN_c \mp i_aN_a \qquad \text{for core 2}$$

For the analogue representation therefore this configuration must be made of two independent transductors of the type considered in Section 2.2 suitably interconnected. It is then necessary to provide a control m.m.f. of $+i_cN_c$ to one core and $-i_cN_c$ to the second. The control current in this case is affected by transformer action from both cores and the effect of each must be superimposed. The analogue of the symmetrical arrangement is shown in Fig. 8.

For zero control signal e_c the magnitude of e_a is adjusted until the cores just saturate, i.e.

$$H_{\max} = H_S = \frac{1}{N_a}\left(\frac{\hat{\imath}_a}{R_L}\right)$$

Examples of waveforms obtained with this configuration are shown in Fig. 9 and the variation of load current with control current, i.e. the 'V' curves are shown in the following section.

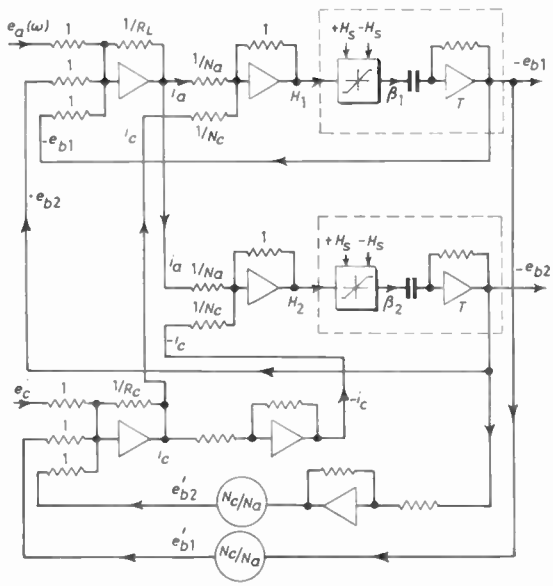
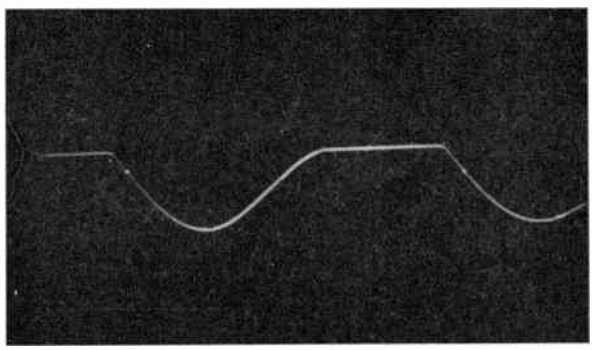
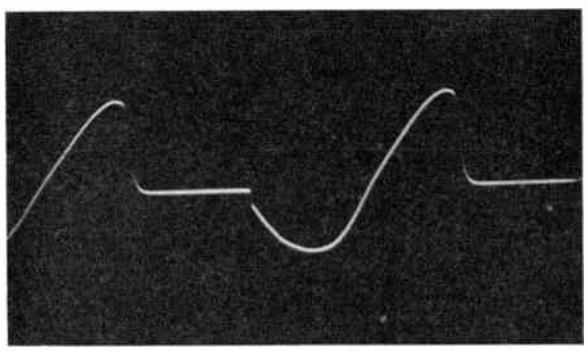


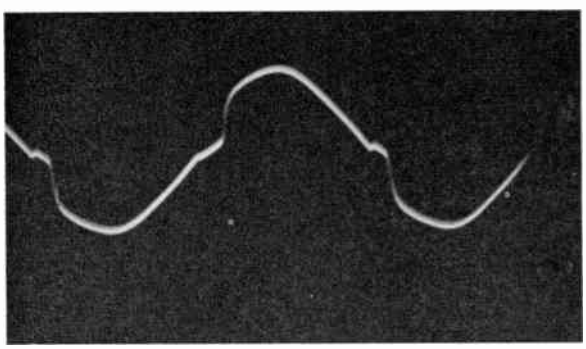
Fig. 8. Analogue of full-wave transductor.



(*a*) Flux.



(*b*) Voltage across one reactor.



(*c*) Load current.

Fig. 9. Waveforms from analogue of full-wave transductor.

3. The Self-excited Transductor or Magnetic Amplifier

A considerable improvement in gain for the basic full-wave transductor can be achieved by suitable application of positive feedback. Here the load current is rectified and made to flow through additional feedback windings as shown in Fig. 10. The current gain of this configuration can be derived from the m.m.f. equality

$$i_c N_c = i_a N_a - i_a N_f$$

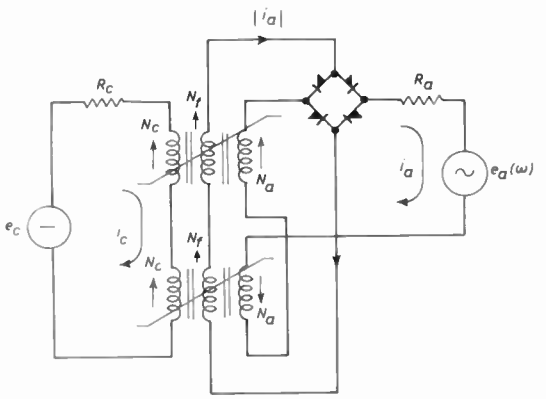


Fig. 10. Self-excited transductor.

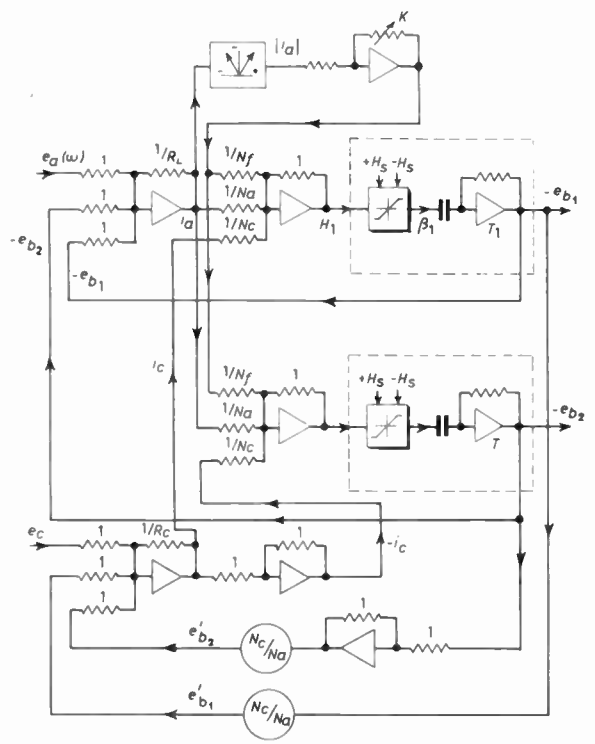


Fig. 11. Analogue of series-connected self-excited magnetic amplifier.

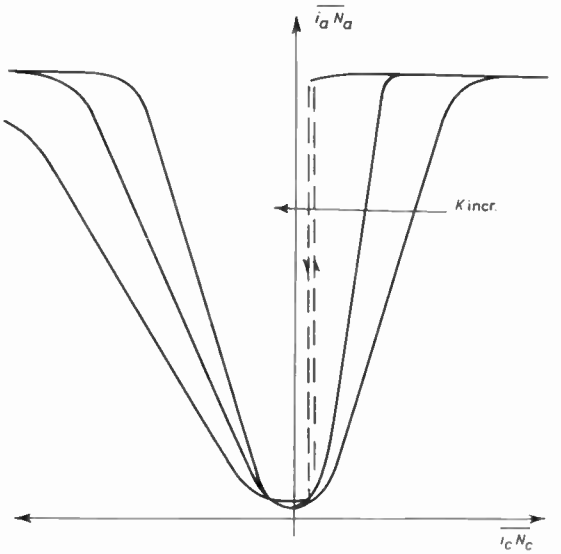


Fig. 12. Variation of load current with control current.

Thus for a core possessing infinite permeability

$$G_i = \frac{\delta i_a}{\delta i_c} = \frac{N_c}{N_a - N_f} = \frac{N_c/N_a}{1-K}$$

which tends to infinity when $N_a = N_f$, i.e. $K = 1$ or 100% self-excitation.

To the analogue of the full-wave transductor of Fig. 8 additional m.m.f.'s of $|i_a|N_f$ are therefore necessary. A modulus unit is thus used to 'rectify' the load current and a proportion K of this is then fed to the cores. The complete analogue of the series-connected self-excited magnetic amplifier is then as shown in Fig. 11. By suitable additions to this network the effect of fixed bias to the cores can be shown.

Figure 12 shows the variation of load m.m.f. $i_a N_a$ against control m.m.f. $i_c N_c$ for a range of feedback factors including more than 100% self-excitation. This results in the characteristic 'jump' (or trigger) and hysteretic phenomena associated with these conditions.

4. General

Conventional d.c. computing amplifiers may be used in synthesizing the analogues of the various configurations. Since operation is repetitive, e.g. 50, 400 or 1600 c/s, all but the m.m.f.-summing amplifiers need be d.c.-coupled to the following unit.

A medium-gain (~ 2000) one-stage amplifier has been developed[4] for such applications. This has a starved long-tail input pair and cathode follower output for low output impedance. Bootstrap feedback is taken from the output to the anode resistance of the first valve. Further a large feed-forward interstage

capacitor ensures that the open loop gain of the amplifier reaches its peak value at a low frequency. The amplifier has a gain of approximately 200 at zero frequency and approximately 2000 over the range 10 c/s–5 kc/s, sufficient for a fundamental frequency of 50 c/s.

The non-linear units are of the conventional diode-gate type.

5. Acknowledgment

The author is indebted to Messrs. J. T. Connolly and H. M. Toner, formerly of the Electrical Engineering Department, Queen's University, Belfast, for the experimental work in this project.

6. References

1. E. H. Frost-Smith, "The Theory and Design of Magnetic Amplifiers", (Chapman & Hall, London, 1958).
2. A. G. Milnes, "Transductors and Magnetic Amplifiers", (Macmillan, London, 1957).
3. H. F. Storm, "Magnetic Amplifiers", (Wiley, New York, 1955).
4. R. E. King, "A medium-gain bootstrap amplifier for repetitive analogue computation", *Electronic Engineering* (to be published).

Manuscript received by the Institution on 10th February 1964. (Paper No. 918.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **June 1964**

June 1964	GBR 16kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	June 1964	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	— 151·6	— 149·0	+ 4	16	— 150·4	— 151·2	— 4
2	— 149·3	— 149·9	+ 6	17	— 151·6	— 151·8	— 4
3	— 149·6	— 150·4	+ 6	18	— 150·7	— 150·7	— 4
4	— 150·7	— 151·8	+ 6	19	— 151·2	— 150·2	— 3
5	— 151·2	— 151·1	+ 4	20	— 150·3	—	— 2
6	— 150·3	— 150·4	+ 5	21	— 150·1	— 150·2	— 1
7	— 150·5	— 151·1	—	22	— 150·4	— 151·2	— 1
8	— 151·3	— 150·2	— 6	23	— 150·8	— 151·4	0
9	— 150·5	— 150·4	— 5	24	— 150·9	— 152·0	— 1
10	— 150·5	— 150·5	— 5	25	— 151·1	— 149·9	0
11	— 151·2	— 150·1	— 5	26	— 149·7	— 150·6	— 1
12	— 150·5	— 151·1	— 5	27	— 149·9	— 150·7	—
13	— 151·1	— 151·1	— 3	28	— 150·0	— 151·1	—
14	— 150·6	— 150·7	— 4	29	— 149·5	— 150·5	— 1
15	— 149·6	— 150·9	— 4	30	— 150·4	— 150·6	+ 1

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

U.D.C. 621.396.677.45

Series and Parallel-fed Linearly Polarized Helical Aerials

By

R. A. CLARK†

AND

T. S. M. MACLEAN†

Summary: A comparison is made between series-fed and parallel-fed linearly polarized helical aerials. The series-fed aerial has been found to have a polarization which rotates significantly with frequency, and the magnitude of this rotation has been found to be in good agreement with the theoretical predictions of the sheath helix analysis. The parallel-fed aerial is useful over the same bandwidth as the single helix for spacings greater than half a wavelength.

1. Introduction

Although the axial-mode helical aerial is relatively a broadband device it suffers from the disadvantage in some applications of being circularly polarized. Hence its power gain with respect to a linearly polarized source is only one half that of some other end-fire aerials of equal length, such as the Yagi. Several attempts have been made to develop linearly polarized versions of the helix, but those of simplest form, namely the zig-zag aerial,[1] and the contra-wound helical aerial,[2] both have bandwidths much smaller than that of the single helix. The Saucisson aerial,[3] formed from a helix wound round a two-wire transmission line, has a wide bandwidth, but in its present form is a long aerial requiring a supporting structure. The purpose of this paper is to present results for two other linearly polarized helices in the form of series- and parallel-fed arrangements. Both were originally proposed by Kraus,[4] and a few experimental results for the parallel case have been given by Jones,[5] but the comparative results presented here show that there is a wide difference in performance between the two cases.

2. The Series-fed Linearly Polarized Helical Aerial

This aerial consists of two oppositely wound helices of equal diameter and pitch angle connected in series (Fig. 1). Linear polarization is achieved as the resultant of two circularly polarized travelling waves having opposite directions of rotation.

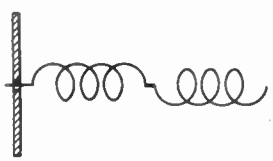


Fig. 1. The series-fed linearly polarized helical aerial.

In the case of the experiments reported here the mean diameter of the helices used was 0·320 in (8·13 mm) and their pitch angle was 13° 40′. The ground plane was 1¼ in (31·75 mm) square, corresponding to approximately one wavelength at the X-band frequencies of operation.

2.1. *Experimental Results*

2.1.1. Radiation patterns

The series-fed helix under test was used as a vertically polarized receiving aerial and radiation patterns were recorded in both the E and H planes. Figure 2

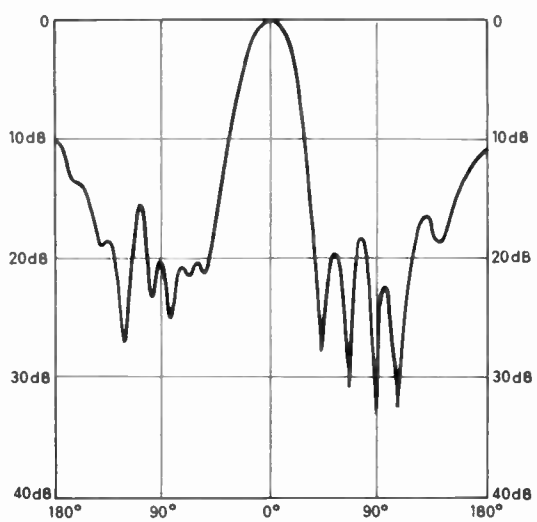


Fig. 2. Radiation pattern of 6 + 6 aerial. $ka = 0{\cdot}723$.

is the H-plane radiation pattern obtained at 8·5 kMc/s for an aerial with six turns in each section. Patterns in the other plane were closely similar. For the same aerial, Fig. 3 shows the first sidelobe level as a function of frequency together with similar results for an aerial with six turns in the inner section and twelve turns in the outer, constructed for reasons given in section 2.1.2.

† Department of Electronic and Electrical Engineering, University of Birmingham.

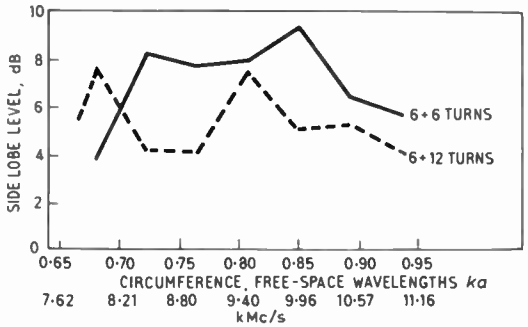


Fig. 3. Sidelobe level vs frequency.

Radiation patterns were also recorded for an aerial with five turns in each section and these were compared with patterns for a ten-turn single helix. The positions of the first and second nulls in these patterns are shown in Fig. 4. The beamwidths of the two aerials are seen to be virtually equal but the power gain with respect to a linearly polarized source, is twice as great for the series-fed helix as for the single helix of equal length.

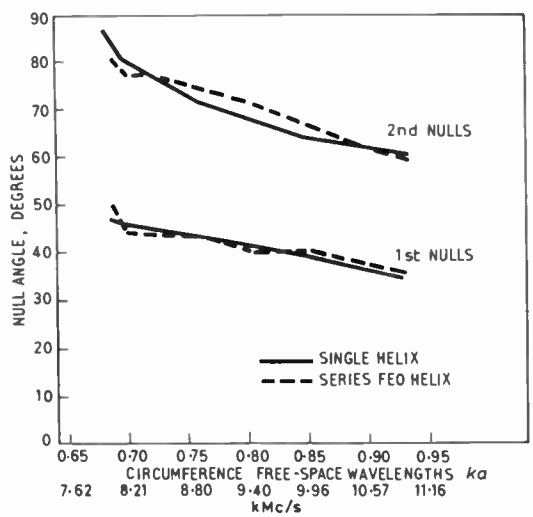


Fig. 4. Angles of nulls in radiation patterns.

2.1.2. Axial ratio

The axial ratio was measured by recording the signal from the helix under test as the transmitting horn was rotated about its axis. The results for the six-plus-six turn aerial are shown in Fig. 5, where the axial ratio is seen to be rather low. It was thought that an improvement might be obtained by increasing the number of turns on the outer section, because a wave travelling from the inner section to the outer has to change from its original direction of rotation to the opposite direction, and this change may be distributed along the outer section. Consequently, the effective number of turns in each section will not be equal. The two other curves in Fig. 5 show that an increase in axial ratio can be obtained in this way, though the improvement is dependent upon frequency as well as on the number of turns.

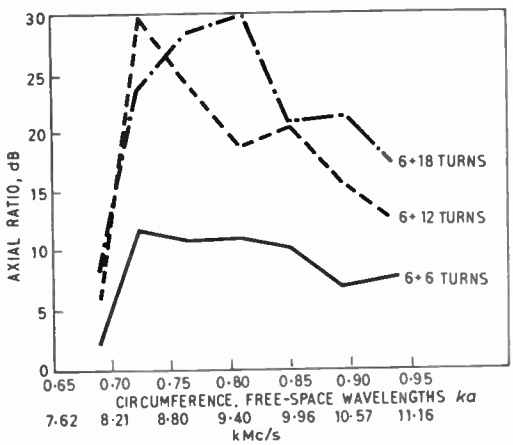


Fig. 5. Axial ratio vs frequency.

The axial ratio ($A.R.$) can also be expressed theoretically in terms of the amplitudes of the oppositely polarized signals, as shown by Kraus[4]

$$A.R. = \frac{E_1 + E_2}{E_1 - E_2}$$

where E_1 and E_2 are the amplitudes of the oppositely rotating vectors. This gives

$$\frac{E_1}{E_2} = \frac{A.R. + 1}{A.R. - 1}$$

This quantity is shown in Fig. 6 for an aerial with four turns in the inner section and up to eighteen in the outer section. It is seen that the turns ratio for exactly linear polarization, when $E_1 = E_2$, increases rapidly

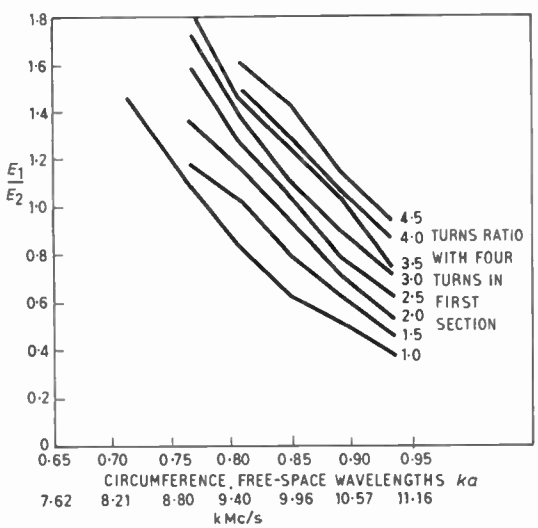


Fig. 6. Ratio of oppositely polarized signals vs frequency.

with frequency when the experimental values of axial ratio are used. Similar results were obtained for the aerial with six turns on the inner section but the variation of turns ratio with frequency is less marked, and it has not been possible to predict this ratio theoretically. This variation is a disadvantage of this form of aerial, but there is a greater disadvantage as shown in the following paragraph.

2.1.3. Plane of polarization

As the axial ratio measurements were being carried out over a wide frequency range it was observed that the plane of polarization rotated. The angle of rotation from the position measured at the lower cut-off frequency is shown in Fig. 7 for the six-plus-six turn

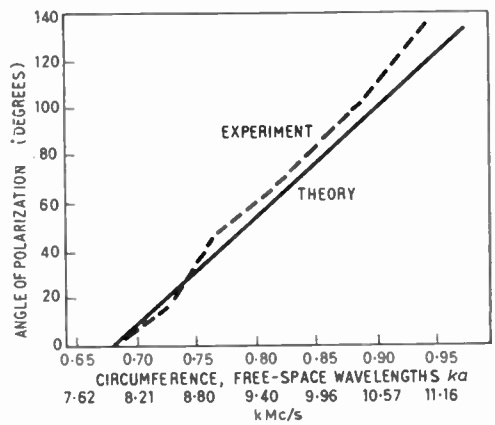


Fig. 7. Rotation of plane of polarization vs frequency.

aerial where it is seen that if this aerial is orientated for maximum signal when the circumference is 0·68 wavelengths, it will receive very little signal when the frequency is increased till the circumference is 0·86 wavelengths. For infinite axial ratio there would be no signal at all.

2.2. *Theoretical Consideration of the Plane of Polarization*

Considering the operation of the aerial as a transmitter, it will be assumed that a travelling wave is produced which follows the direction of the conductor along the first section of the aerial. As it continues along the second section an increasing proportion of its energy follows the new conductor direction. The energy which does not follow a conductor corresponds to the -1 mode of the sheath helix analysis,[6] which has a much lower phase velocity than the normal $+1$ mode along the conductor in the operating frequency range. The direction of the plane of polarization is mainly determined by the large phase delay of this -1 wave along the second half of the aerial, as will now be shown.

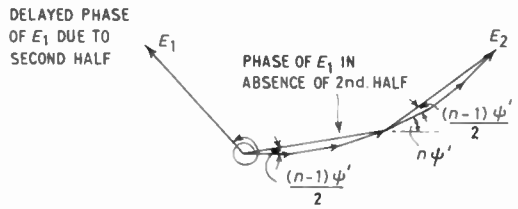


Fig. 8. Relative time phase of fields of oppositely polarized waves at a distant point.

Figure 8 shows the relative phase of the resultant fields, E_1 and E_2, due to the waves following the conductor in each section at a distant point on the axis of the aerial. As a first approximation, the fields due to individual turns in the second section are assumed to be equal, and it is also assumed that the radiation from the wave not following a conductor is negligible. The phase difference between the fields due to adjacent turns at a distant axial point is given by

$$\psi' = 2\pi \cdot p_\lambda \cdot \left(1 - \frac{c}{v}\right)$$

where p_λ = pitch in free-space wavelengths, v = axial phase velocity. The values of ψ' are small for the wave following a conductor since its axial phase velocity is almost equal to the velocity of light. However, the phase of E_1, with respect to the field due to the first turn, must be increased by an amount equal to the phase delay along the second section of n turns with the original direction of rotation. This additional phase delay is given by

$$n \cdot 2\pi p_\lambda \frac{c}{v_{-1}}$$

where v_{-1} is taken as the axial phase velocity of the -1 mode in the sheath helix analysis.[6]. Since this phase velocity is a function of frequency so also is the phase delay and hence E_1 rotates with frequency. Numerical values of the phase delay for a six-plus-six turn aerial are more than 360 deg.

Figure 9 shows the effect of a change of frequency upon the instantaneous direction of polarization of E_1, and E_2 at a distant point and upon the resultant direction of linear polarization. The theoretical curve in Fig. 7 shows the angle of rotation of the plane of polarization calculated for a six-plus-six aerial, and in

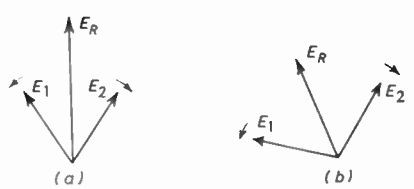


Fig. 9. Effect of change of frequency upon the position of the plane of polarization.
(*a*) $ka = 0{\cdot}68$. (*b*) $ka = 0{\cdot}74$.

view of the assumptions made the experimental agreement is surprisingly good.

3. The Parallel-fed Linearly Polarized Helical Aerial

This aerial consists of two oppositely wound helices connected in parallel to form a two-element broadside array (Fig. 10). Details of the dimensions of the

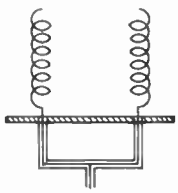


Fig. 10. The parallel-fed linearly polarized helical aerial.

helices used are given in Table 1, two diameters being required to determine the lower and upper cut-off frequencies with the available equipment.

Table 1

Aerial	Mean diameter $2a$	Pitch angle	Frequency for which $ka = 1$
1	0·320 in	13° 40′	11·77 kMc/s
2	0·496 in	13° 15′	7·58 kMc/s

The helices were ten turns in length and were mounted on a rectangular ground plane, one wavelength wide by four wavelengths long. Measurements were made on aerials with spacings varying from 0·4 to 1·5 wavelengths between the helices.

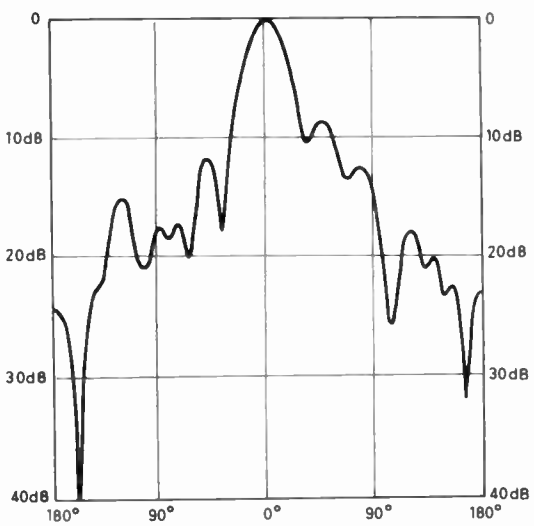


Fig. 11. Radiation pattern of aerial 1.
$ka = 0{\cdot}723$. Spacing = 2 cm = $0{\cdot}5\lambda$.

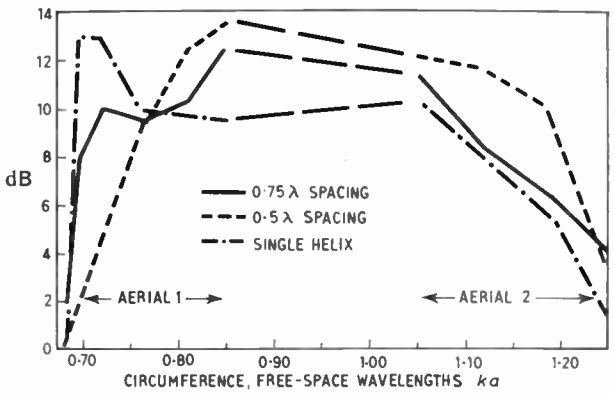


Fig. 13. Sidelobe level vs frequency.

3.1. *Experimental Results*

3.1.1. Radiation patterns

The parallel helices were mounted in the horizontal plane and radiation patterns in the plane of the array were recorded for various spacings using a vertically polarized source. The polarization of the array was made vertical by ensuring that the start leads of the individual helices were vertical. Typical patterns obtained at two different spacings are shown in Figs. 11 and 12. Figure 13 shows the difference in level between the main lobe and largest sidelobe plotted against the circumference in free-space wavelengths, ka, for two spacings and also for a single helix. The bandwidth between the points where the sidelobe is 6 dB is seen to be practically constant for the cases shown, though the cut-off frequencies at 0·5 wavelength spacing are approximately 5% above those for the single helix.

For spacings of less than 0·5 wavelength it was found that the radiation patterns obtained at frequencies

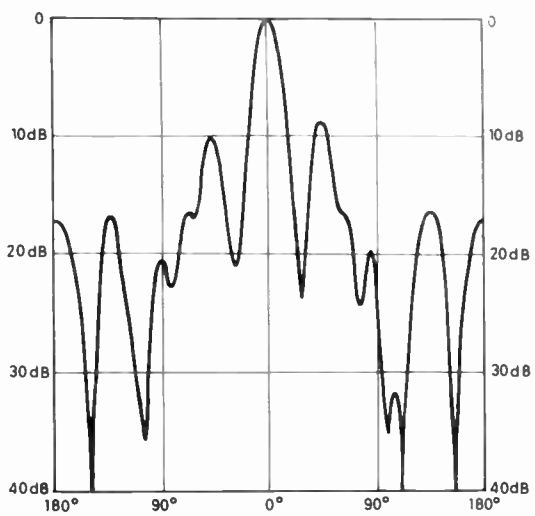


Fig. 12. Radiation pattern of aerial 2.
$ka = 1{\cdot}056$. Spacing = 4 cm = $1{\cdot}0\lambda$.

below $ka = 0{\cdot}88$, were broader than those characteristic of end-fire aerials operating with increased directivity and had sidelobe levels as low as -3 dB.

3.1.2. Axial ratio

The variation of axial ratio over the operating frequency range of the aerial is shown in Fig. 14 for a spacing of 0·75 wavelength. Similar results were obtained for other spacings, the lowest value of axial ratio being 12 dB which is equivalent to a polarization ellipse with a major-to-minor axis ratio of approximately 4 : 1.

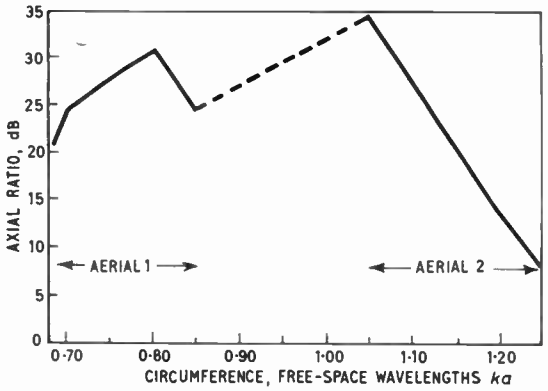


Fig. 14. Axial ratio vs frequency. 0·75λ spacing.

3.2. *Theoretical Considerations*

In general, the radiation pattern of an array is given by the product of the pattern of an individual element and the appropriate array factor. This is applicable in this case, even though oppositely polarized helices are used, since the individual patterns are independent of the direction of rotation of the polarization. The directivity of the parallel-fed helical aerial can therefore be varied by adjusting the length of the helices and the distance between them.

However, pattern multiplication is only applicable when the individual patterns are unaltered by placing the helices in close proximity. In order to determine the spacing at which coupling between parallel helices becomes significant the power distribution round a helix has been calculated for the $+1$ mode from results given in the sheath helix analysis of the helical aerial.[6] From this the power density due to one helix at the surface of a second helix may be determined, and the ratio of this quantity to the power density at the second helix due to its own excitation gives a measure of the coupling between the helices. This ratio is plotted against spacing in Fig. 15 and it is seen that its value increases rapidly as the spacing is reduced from 0·5 wavelength.

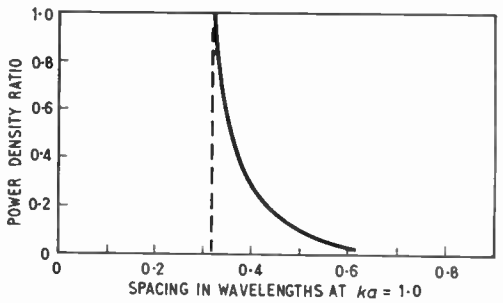


Fig. 15. Relative power density at surfaces of parallel helices.

It has been found that the bandwidth of the contra-wound helical aerial[2] is much less than that of a single helix, and it may be expected that the bandwidth of the parallel-fed helical aerial is reduced as the spacing between the axes of the helices is reduced below 0·5 wavelength. For spacings less than 0·32 wavelength at $ka = 1$, however, constructional difficulties occur, since the helices become partly interleaved.

4. Conclusions

Results of measurements on the series- and parallel-fed linearly polarized helical aerials have been presented. The series-fed aerial, though capable of being made linearly polarized over a useful frequency range, has the characteristic that its plane of polarization rotates with frequency. The bandwidth of the parallel-fed helical aerial has been shown to be the same as that of the circularly polarized single helix for spacings between the helices of at least 0·5 wavelength.

5. References

1. W. A. Cumming, "A non-resonant endfire array for v.h.f. and s.h.f.", *Trans. Inst. Radio Engrs* (*Antennas and Propagation*), **AP–3**, p. 55, 1955.
2. R. A. Clark and T. S. M. Maclean, "The contra-wound linearly polarized helical aerial", *Electronic Engng*, **36**, pp. 80–3, February 1964.
3. E. Spitz, "A class of a new type of broadband antenna" in "Electromagnetic Theory and Antennas", edited by E. C. Jordan, Part 2, pp. 1139–48 (Pergamon Press, Oxford, 1963).
4. J. D. Kraus, "Antennas" (McGraw-Hill, New York, 1950).
5. G. C. Jones, "An experimental design study of some S- and X-band helical aerial systems", *Proc. Instn Elect. Engrs*, **103B**, pp. 764–71, November 1956. (I.E.E. Paper No. 2170R, November 1956.)
6. T. S. M. Maclean and W. E. J. Farvis, "The sheath-helix approach to the helical aerial", *Proc. Instn Elect. Engrs*, **109C**, pp. 548–55, 1962. (I.E.E. Monograph No. 519E, May 1962.)

Manuscript first received by the Institution on 6th January 1964, and in final form on 22nd April 1964. (*Paper No. 919.*)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

VIBRATING CAPACITOR

A new vibrating capacitor has been developed by Dutch engineers for application in electrometers. A thin glass membrane is clamped between two glass insulators, the middle section of which is hollow-ground. These middle sections and both faces of the membrane are coated with tantalum, thus constituting two capacitors, the capacitance of which varies periodically when the glass membrane vibrates. One is the actual vibrating capacitor, the other serves as the capacitive drive of the membrane and forms part of a 1-Mc/s transistor oscillator. This h.f. voltage is amplitude modulated at the natural frequency of the membrane (6 kc/s). Since the frequencies of the a.m. voltage are much higher than that of the alternating voltage into which the measured d.c. signal is converted, no interference occurs. Variation of ambient temperature causes a zero-point drift of no more than about 15μV/deg C.

"A vibrating capacitor driven by a high-frequency electric field", A. G. van Nie and J. J. Zaalberg van Zelst. *Philips Technical Review*, **25**, No. 4, pp. 95–103, 1963/64.

ELECTROLYTIC TANK INVESTIGATIONS ON C.R.T. GUNS

Electron path tracing equipment, based on the electrolytic tank analogue of the electrostatic field, has been used in Australia to trace electron paths through an electron gun of a design commonly used in modern television picture tubes. Electron paths were traced in the three parts of the gun (the cathode system, the accelerating system and the final electrostatic focusing lens) under various conditions. The equipment can be used to illustrate the main features of the action of the electron gun in forming a fine spot on the screen. It was found necessary to support the tank measurements by using Leibmann's measurements of spherical aberration of a unipotential lens in explaining the growth of spot size with drive.

"An investigation of a television picture-tube gun", D. M. Sutherland. *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **25**, No. 2, pp. 99–112, February 1964.

COMPENSATION OF TRANSMISSION ERRORS IN THE PAL COLOUR TELEVISION SYSTEM

Transmission errors, which may cause colour distortion in pure N.T.S.C. colour television systems, can be compensated in the PAL systems by suitable change-over of one modulation axis from one line to the next by employing an ultrasonic delay line in the receiver. A German engineer has shown that the separation of the two colour difference signals, which requires phase-sensitive synchronous detectors in the case of the N.T.S.C. system, is achieved in the PAL system by in-phase addition and subtraction of the signals from two lines prior to demodulation. Colour cross-talk due to quadrature errors is compensated in this case. In a modified PAL receiver the reference carrier-to-carrier injection is derived directly from the signal. This will remain in synchronism with the signal and avoid colour distortion even when the tape speed of a television recorder changes.

"The PAL television system—principles of modulation and demodulation", W. Bruch. *Nachrichtentechnische Zeitschrift*, **17**, No. 3, pp. 109–21, March 1964.

LOW-NOISE RECEIVER FOR C.S.I.R.O. RADIO TELESCOPE

The C.S.I.R.O. 210-ft telescope at Parkes, New South Wales, has a beamwidth of 7·5 minutes of arc and is potentially capable of distinguishing 100 000 radio sources at a wavelength of 11 cm. in the area of the sky accessible to it. However the majority of sources are extremely weak and a highly sensitive receiver is required to detect them. With sources present in large numbers it is desirable to achieve the required sensitivity without recourse to long integration times. A step toward achieving the required sensitivity has been made through the use of a broadband receiver in conjunction with a low-noise varactor diode parametric pre-amplifier, used to lower the system noise temperature to less than 100° K.

"A low-noise 11-cm receiver installation on the C.S.I.R.O. 210-ft radio telescope", B. F. C. Cooper, T. E. Cousins and L. Fruner. *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **25**, No. 4, pp. 221–27, April 1964.

SEMICONDUCTOR NUCLEAR PARTICLE DETECTORS

In the past few years, the use of semiconductor particle detectors has been extended to beta- and gamma-ray spectroscopy as well as to higher energy nuclear reaction studies. This is apparently due to the availability of deep drifted surface passivation. Particle identification has been facilitated in France by the development of thin ΔE detectors. The characteristics of various types of detectors, their rise-time and carrier collection efficiency can be deduced from the operating principles which have been discussed. The response of nuclear particles, including beta and gamma rays, illustrates the performance of semiconductor detectors. Major areas of investigation include irradiation effects and the development of new materials for gamma spectroscopy.

"Semiconductor nuclear-particle detectors", J. W. Mayer. *L'Onde Électrique*, **44**, pp. 342–54, April 1964.