THE RADIO AND ELECTRONIC ENGINEER

The Journal of the British Institution of 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 26 DECEMBER 1963 NUMBER 6

FAIR EXCHANGE

By the will of its members the Institution has submitted to Her Majesty's Privy Council a request to change the name of the Institution to 'The Institution of Electronic and Radio Engineers'. The purpose is to recognize the fact that the art and science of radio engineering has been encompassed by the more comprehensive scope of electronic engineering.

The increasingly diverse applications of the thermionic valve have been regularly recorded in papers published in the Institution's *Journal*, and the spectacular development of semiconductor devices has given further impetus to the extension of radio engineering into wide fields. In fact, a very large proportion of the members of the Institution are now concerned with matters other than 'radio'—or aural and visual methods of communication. This was recognized by the adoption of the title *The Radio and Electronic Engineer* for the *Journal*.

Both the Institution's name and that of its *Journal* now fairly describe the nature of the Institution's work which, however, wholly depends on the willingness of members and readers to indulge in fair exchange. The Institution plays its part in providing the forum and an opportunity for its members and other readers of the *Journal* to benefit from its publications. An essential part of exchange, however, is the willingness of members and readers to contribute. Indeed, one of the obligations of an engineer is to further his profession and the contribution of papers is one of the ways in which this may most worthily be effected.

Many of the papers published in the Institution's *Journal* have first been presented at an Institution meeting. The suitability of a paper for oral presentation is not, however, an essential requirement and papers for publication only are just as welcome as papers for reading before a meeting.

The merit of papers published in *The Radio and Electronic Engineer* is reviewed each year by the Council for the granting of Premiums and Awards. These Premiums and Awards relate to nearly all the varied activities of the radio and electronic engineer, and are regarded throughout the profession as authoritative recognition of outstanding work.

Because it is a leading journal in the science and technology of radio and electronics, the world-wide circulation of *The Radio and Electronic Engineer* is augmented by inclusion of all papers published in it in the main abstracting journals of the world. Thus it provides an unrivalled opportunity for bringing new work to the notice of other scientists and engineers concerned with the same or allied tasks. This direct and indirect circulation is essentially beneficial in both directions. Many authors are able subsequently to discuss their work informally by correspondence with others engaged on similar problems; in this way the risk of embarking on the solution of questions already answered is greatly reduced.

For these and many other reasons all engineers are urged to consider publication of papers on their work as soon as it reaches a suitable stage, and thereby contribute to promoting the fair exchange of ideas. The Papers Committee welcomes contributions from engineers all over the world, irrespective of nationality or membership of the Institution, and is always pleased to advise on the best form of presentation.

F. W. S.

INSTITUTION NOTICES

The Institution's New Title

At a Special General Meeting of Corporate Members, held in London on 27th November last under the chairmanship of the President, Admiral of the Fleet the Earl Mountbatten of Burma, K.G., a resolution of the Council, that the title of the Institution should be changed to "The Institution of Electronic and Radio Engineers" was passed by an overwhelming majority which included over 600 proxy votes in favour. The Resolutions were as printed in the November issue of *The Radio and Electronic Engineer*. The Resolutions have now been submitted to The Lords of Her Majesty's Privy Council for formal approval to the amendment to the Royal Charter of Incorporation and to the appropriate Bye-Laws.

A detailed report of the Special General Meeting, and also of the Annual General Meeting which followed it, will be published in the January 1964 issue of the *Proceedings*; an abridged report will appear in *The Radio and Electronic Engineer* for January. The meetings on 27th November were followed by the Presidential Address of Mr. J. L. Thompson and this will also be published in the January issue of *The Radio and Electronic Engineer*.

I.F.A.C. 'Report Back' Conference

Following the Second International Congress of the International Federation of Automatic Control in Basle from 28th August to 4th September 1963, the United Kingdom Automation Council is organizing a 'Report Back' Conference in Manchester from 7th to 9th January 1964. More than 150 papers were presented by international experts, surveying the state of automation in various fields. The 'Report Back' Conference is being held to examine and appraise the Basle proceedings and to study the important aspects in greater detail. A team of rapporteurs, all of whom took part in the Basle Conference, will present synopses of the various sessions. Early application is invited and should be addressed to The Secretary, The Institution of Production Engineers, 10 Chesterfield Street, London, W.1.

International Conference on Instrumentation and Measurement

The third International Measurement Conference ('IMEKO') is to be held jointly with the sixth International Instruments and Measurements Conference in Stockholm from 14th to 19th September 1964.

The papers, which will probably total 100, will be delivered in four general sections covering such subjects as metrology, instrument design, and instrument electronics and in six special sections which will deal with mechanical, electrical, thermodynamic, physico-chemical and other measurements. Problems lying between 'measurement' and 'automation' will be dealt with in a special section organized in conjunction with International Federation of Automatic Control (IFAC).

Two notable papers selected for presentation at the opening plenary session will be a general review of astronautical instrumentation by Professor N. I. Christjakov (U.S.S.R.) and a survey of measurement standards in the space age by Dr. W. A. Wildhack (U.S. National Bureau of Standards).

All papers will be preprinted and only up-to-date summaries will be presented at the conference. The proceedings will be simultaneously translated into English, French, German and Russian.

The British contribution of the Conference is being sponsored by the United Kingdom Automation Council (formerly the B.C.A.C.), secretarial functions being carried out by the Society of Instrument Technology. The Brit.I.R.E. representative on the British Organizing Committee is Mr. A. G. Wray, M.A. (Member), a member of the Institution's Council and chairman of its Programme and Papers Committee.

Further information regarding the conference may be obtained from The IMEKO Secretariat, P.O.B. 3, Budapest 5, Hungary, *or* The Swedish Conference Committee I. & M. : I.V.A., P.O.B. 5073, Stockholm 5, *or* The Secretary, Society of Instrument Technology, 20 Peel Street, London W.8.

President of International Council of Scientific Unions

At the Tenth General Assembly of the International Council of Scientific Unions held in November in Vienna, Dr. H. W. Thompson, C.B.E., F.R.S., of St. John's College, Oxford, was elected President of the Council for the next two years. Professor D. Blaskovic of Czechoslovakia was elected Secretary-General and Ing. Gén. G. Laclavère of France was re-elected as Treasurer.

Correction

The following amendments should be made to the paper "A Series Type D.C. Negative Resistance for Analogue Computers" which was published in the November issue of *The Radio and Electronic Engineer*.

Page 420: Section 9. References—in references 5 and 7, *J. Appl. Phys.* should read *Brit. J. Appl. Phys.*

Completion of Volume 26

This issue completes Volume 26 of the *Journal* which covers the period July–December 1963. An index to the volume will be circulated with the January issue.

U.D.C. 615.477: 616.74

Muscle Substitutes and Myo-Electric Control

By

A. BOTTOMLEY, M.B., B.Chir.†

A. B. KINNIER WILSON, M.B., M.R.C.P.‡

AND

A. NIGHTINGALE, Ph.D.§

Presented at the Symposium on "Practical Electronic Aids for the Handicapped" in London on 28th March 1962.

Summary: Muscle substitutes are defined and a description is given of some of those in use and under development. The possibility is suggested of controlling these devices by using the electrical potentials picked up over contracting muscles. Past experimental work in the field is discussed and the problems associated with it are described. Finally, the authors' experimental control system is described and its performance is related to these problems.

1. Muscle Substitutes

Muscle substitutes are designed to actuate prostheses (replacements for amputated parts of the body) and ortheses (aids for limbs which have been paralysed). Characteristically they make use of an external source of energy, i.e. the energy is not supplied by some other part of the patient's musculature. Two classes of energy source have so far been employed, electrical and pneumatic. Papers on the relative merits of the two systems, and of alternative gases for pneumatic power, are to be found in the Report on the Lake Arrowhead Conference, 1960.[1] Electrical energy has been used in the U.S.A. by Alderson,[2] in Russia by Kobrinskii and his co-workers,[3] and in France, where a development of the so-called Vaduz electric hand is made.[1]

When one considers the problem of providing myo-electric control, there are many advantages in using an electrical actuator as a component in what is essentially an electrical servo system. However, electric actuators have to be powered by small electric motors running at high speeds with gearing, and these are usually somewhat heavy and noisy, and give rise to vibration. Storage of electrical energy is also a problem. Further, three movements are often required in an arm—elbow bending, wrist turning and prehension (grasping)—and to provide three electric motors capable of the required torque when geared down, coupled with an adequate turning/bending speed of about 12 rev/min, would mean that with the motors available at present the arm would become very heavy.

The consensus of opinion in the U.S.A. and Europe seems to be that pneumatic energy is to be preferred. Liquid carbon dioxide is readily available and can be carried in light-weight containers. Pneumatic actuators can be made to operate smoothly and are relatively quiet and light. Pneumatic energy derived from compressed carbon dioxide has been used in the Heidelberg arm,[4, 5] the McKibben muscle,[6, 7] the American Institute for Prosthetic Research arm,[8] and the Hendon pneumatic prosthesis motor.[9]

There have been several different mechanical ways of providing pneumatic units for artificial arms and prostheses, and they will be described in increasing order of sophistication. The first one is the McKibben muscle, which consists of a rubber tube covered with a nylon braid. When gas is admitted to the muscle the rubber tube bulges; because of the nylon braid it also shortens, thus producing a pull. This pull is not very strong nor linear and the unit shows considerable hysteresis when contraction and relaxation are compared. It is therefore largely confined to providing a grasping action by motorizing light-weight splints for the finger and thumb of patients who require extremely light-weight units: for this purpose the McKibben muscle is admirable.

For all other purposes pistons in cylinders or bellows are used. (Vane motors are probably more expensive in gas and certainly make quite a loud noise.) There are two factors in considering piston or bellows motors. One is that the movement must be restrained otherwise an uncontrolled movement from one extreme position to the other would occur. This can be done either by making the motor pull against a spring, or pull against another piston, the movement of which is controlled. Piston movement against a spring is similar to that which occurs when a pair of bellows is used, in as much as there must be a spring return. Movement against a piston can be done in one of two ways. Either the cylinder is double ended and the piston has two-way seals, or a separate cylinder is used with a separate piston. It is here that the second factor comes in. The even pull of the piston should result in output which also has an even torque. This is best done by means of a pulley or

† St. Thomas' Hospital, London, S.E.1.

‡ Poliomyelitis Centre, West Hendon Hospital, Middlesex.

§ St. Thomas' Hospital until his death in January 1963.

gearing system and not by means of a rod and lever, or rod and crank, which inevitably must produce a different torque at different positions of the stroke.

The method adopted for the Hendon motor is to provide two separate cylinders with cables coming straight from the pistons (Fig. 1). This avoids the use of piston rods and enables a more compact motor to be made. From one piston a cable runs twice round a pulley to return to the other piston. This provides an even torque at all positions with the minimum amount of wasted effort, in as much as for maximum pull the only thing to be overcome is the friction of the seals. If a spring is used then this always inevitably has to be overcome, and this of course is the case also with bellows.

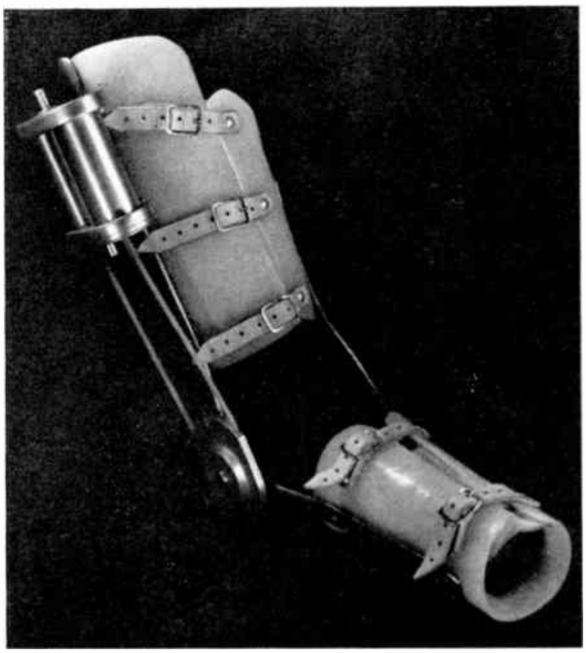


Fig. 1. Hendon motor applied to arm splint.

The most important attribute of a motor for splints or prosthesis is, of course, patient acceptability. Patients with disabilities, amputations or paralysis, develop a considerable ability to get along somehow in spite of their difficulties. This develops so rapidly that in fact should an amputee fail to be provided with an artificial arm within eighteen months of the amputation, he becomes effectively one-armed and will not adopt the prosthesis except for cosmetic reasons. It follows, therefore, that the motor must have two properties *par excellence*. One is high power-to-weight ratio and the other is controllability.

The Hendon motor has been designed with the maximum controllability in view. This means that its output torque must be even throughout the whole range of movement and it must be absolutely smooth. At the same time it seems desirable to provide a measure of acceleration in order to obtain not only a natural looking movement but also one in which the patient can produce a greater effect the harder he pushes the lever. This avoids the difficulty of thinking in terms of switching on and waiting until the arm comes to the right position and then switching off again. This is an unnatural way of thinking, since we naturally think in terms of increased tension, increased pull, increased effort.

As far as controllability is concerned, having achieved a suitable output, the important thing is to make sure that the motor starts without any jerk, continues absolutely smoothly and stops without any bounce. The most elegant way of doing this is by allowing the gas pressure to enter the driving cylinder at full bore without any restrictions and to restrict the outflow of gas from the exhausting cylinder. This gives a much smoother action when compared with the simple control by means of varying the flow of gas into the driving cylinder.

Because of its twin cylinder arrangement, the Hendon motor requires a double valve. Both metal-to-metal and rubber-sealed bobbin valves have been used and at present the metal-to-metal type is preferred as being much smoother in action. Sealing problems are, however, a difficulty and other types are being tried. The valves are generally operated by movements of some other part of the patient's body, e.g. a shrug of the shoulder, or abduction of the knee against a lever (in the case of a patient confined to a wheel chair). This requires a period of learning.

2. Myo-electric Control

It would clearly be an advantage to employ as a control signal the nervous impulses which would normally be used by the patient to control the muscle before the onset of his disability. While it is not practicable to detect motor nerve impulses directly, it is possible to pick up the electrical activity from residual active muscle fibres (in cases of paralysis) or from the muscle stump (in the case of an amputee). This myo-electric activity results from the signals sent along the residual nerve fibres and can thus be used as a measure of the effort being made by the subject to contract that particular muscle.

The possibility of using myo-electric signals to control a powered substitute muscle was under consideration in the U.S.A. as long as 10 years ago, when Berger and Huppert[10] concluded that the method was worth pursuing. At that time the International Business Machines Corporation contained a "Prosthetics Research Division" which was constructing an electrically powered artificial arm, and it was hoped that this might eventually be controlled bio-electrically. However, as far as the present authors are aware, this project did not mature.

In 1955 an experimental myo-electric control system was described by Battye, Nightingale and Whillis,[11] who showed that a simple solenoid-operated split hook could be closed and opened by a patient whose hand had been amputated. The action potentials were picked up by two small surface electrodes placed over the residual muscle of the stump, and the subject was asked to grip with his 'phantom' hand. No attempt was made to construct a practical prosthesis which could be fitted to a patient for daily use, but the experiment provided useful experience, indicating some of the problems which would arise. One was the difficulty of avoiding spurious operation by interfering signals. Mains interference and movement artefacts were reduced by cutting the amplifier band to the range 100 to 1000 c/s, without serious loss of signal. Operation was made more reliable by introducing some 'backlash' into the relay which switched the power to the terminal device, so that a considerable effort had to be made by the subject to close the grip, but the grip could then be maintained with a smaller effort (Fig. 2). The disadvantages of the system were firstly that it was in essence a simple switch, so that grading of the gripping force was not possible, and secondly that only one channel was used so that a conscious relaxation was necessary to release the grip. An essentially similar system (for a paralysed patient rather than an amputee) was reported in 1959 by Geddes, Moore, Spencer and Hoff.[7] These workers used action potentials picked up from forearm muscles to operate a miniature solenoid valve consuming only 1 watt (a performance which cannot be matched by any commercially available device as far as the authors are aware) in which gas flows to and from a McKibben muscle harnessed to the patient's finger and thumb. Again, only one channel was used, although some grading of the output force was possible by controlling the period of *time* during which gas was admitted to the actuator. However, this manoeuvre was not a normal one but had to be learned.

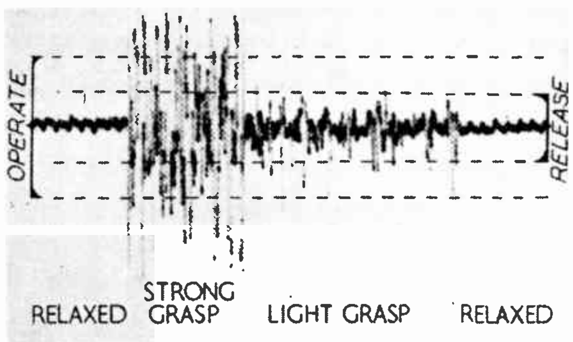


Fig. 2. Wave form of e.m.g. showing amplitude required to 'operate', 'maintain' and 'relax' the terminal device. (From Battye, Nightingale and Whillis[11] by courtesy of the Editor, *Journal of Bone and Joint Surgery*.)

During the last few years a group of Russian workers has reported considerable success with a much more sophisticated system, designed to operate an artificial hand for an amputee (e.g. Kobrinskii *et al.*[3]). Two channels were used, one from the muscles normally closing the hand, the other from those normally opening. In each channel the myo-electric signal was amplified, rectified, and converted into pulses whose repetition frequency was proportional to the original amplitude. The pulses from the two channels were fed in opposition to a 'step by step' electric motor operating the artificial hand as a differential servo. The direction of movement and the force exerted were thus controlled actively, a great improvement over the single channel switch.

Work is in progress at Los Angeles with a view to designing a much more complex system providing natural arm movements by myo-electric control. Weltman and Lyman,[12] describing a preliminary survey of the patterns of myo-electric activity during arm movement, concluded that a control which co-ordinates signals from several muscles through a central logic will be required to provide natural arm movement. They are investigating the effects of special training to produce more distinct patterns of muscular activity, and are considering the possibility of using implanted electrical pick-ups.

Another forward-looking project, recently started at Cleveland, Ohio,[13] also aims at providing natural arm movements. In this case myo-electric control is not envisaged but the system is mentioned here as a possible alternative. A splint is being designed carrying digital transducers indicating motions about axes at the shoulder and elbow. This will be worn by a normal subject who will make a series of natural movements, and the pattern of each movement will be stored on magnetic tape. The patient will wear a similar splint incorporating pneumatic actuators with valves operated by torque motors, controlled by the magnetic store. He will 'call for' a particular movement from the store when he requires it, using some sort of code, possibly using an infra-red transmitter and pick-up behind the object required. This group is also contemplating the possibility of using implanted electrodes with a transmitter to telemeter the muscular activity, and also in reverse to transmit signals into an implanted circuit to stimulate a denervated muscle—an extension of the concept of the r.f. cardiac pacemaker.

3. The Electromyogram as a Control Signal

3.1. *Aim of Present Work*

It was decided that the initial effort should be directed to the paralysed patient although it is of course hoped that the system will also be applicable to amputees. In the first instance, in order to gain experience, the problem was simplified by aiming at helping a patient who still has reasonable control of his hand, but whose upper arm muscles are too weak to enable him to use it. It was felt that the subjects who are paralysed in both arms should be chosen, so that they would have a strong motive to try out any device which might be helpful. Under these conditions a very simple system involving only flexion and extension at the elbow might be useful.

3.2. *The Normal E.M.G. and the Concept of the Motor Unit*

A normal muscle is innervated by a large number of motor nerve fibres. Each nerve fibre branches and innervates some hundreds of muscle fibres, and the functional unit consisting of a single nerve fibre and its associated muscle fibres is called a 'motor unit'.[14] The 'command' is sent down the motor nerve as a small electrical impulse; as this reaches each muscle fibre a similar brief impulse is triggered off in the muscle fibre. This travels along the muscle fibre which a few milliseconds later gives a single mechanical twitch. A sustained contraction is maintained by the twitches of a large number of motor units 'firing' at different frequencies throughout the muscle. Surface electrodes placed on the skin over the muscle pick up an irregular but continuous wave form which is the result of the summation at the electrodes of contributions from a large number of motor units. This wave form is termed an 'interference pattern'. (See Figs. 2 and 3.) Occasionally, during a very weak contraction, it is possible to pick up the activity of a single motor unit, a regularly repeated 'spike'. A normal contraction is graded by the gradual increase in rate of firing of a few motor units and also the gradual recruitment of more motor units as the effort is increased.

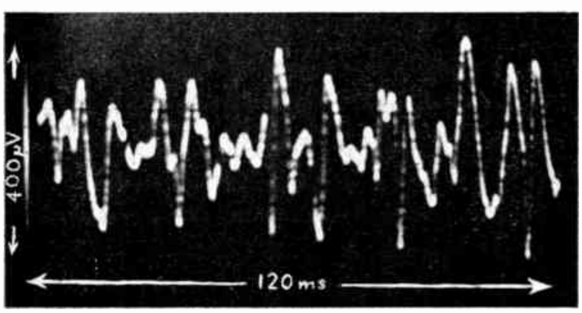


Fig. 3. Electromyogram from a partially paralysed muscle. Patient M.S. Right biceps. Graded 2 on the M.R.C. scale (perceptible movement, but insufficient strength to operate against gravity).

The myo-electric signal is easier to detect than the neuro-electric one (which is usually obscured by tissue and amplifier noise when surface electrodes are used) for the following reasons:

(1) the muscle fibres are larger and produce a greater electrical discharge on excitation;

(2) there are many muscle fibres for each nerve fibre; and

(3) the motor nerves run in bundles close to those for other muscles and to sensory fibres, and it is impossible to separate the signals from the different sources.

3.3. *The E.M.G. from Muscles affected by Poliomyelitis*

A preliminary investigation was carried out on patients whose muscles had been affected by poliomyelitis. Surface electrodes (silver discs 1 cm in diameter) were attached by adhesive strapping to the skin over the muscle, about 5 cm apart, and the subject was asked to attempt a strong contraction. The output from the amplifier was displayed on a cathode-ray tube and photographed. Figure 3 shows the wave form obtained from a muscle which was too weak to operate against gravity. A number of weak muscles in two patients were investigated, and in every case an electrical signal was detected, although in one instance the muscle was graded 'zero,' that is, there were no clinical signs of contraction. The waveform was in most cases fairly continuous and irregular, with an amplitude up to a few hundreds of microvolts (peak to peak). In the weakest muscles the amplitude of the interference pattern was as low as 50 μV, but this could still be comfortably detected by a low-noise amplifier with a balanced input designed to reject in-phase mains interference. In one case of a grade zero muscle large single motor unit potentials were detected.

These findings are in agreement with observations of clinicians who use electromyography to study the development of re-innervation after poliomyelitis. It appears that during re-innervation a particular surviving nerve fibre may connect up to a number of denervated muscle fibres originally belonging to other units. This results in a much smaller final number of motor units, but these on the average contain a larger than normal number of muscle fibres. Buchthal and his co-workers[15] have shown that the motor unit action potential is usually larger in amplitude, and the fibres of a given motor unit are more widely dispersed, in muscles affected by anterior horn cell involvement. Lenman[16] has shown that the e.m.g. is greater in amplitude for a given developed force in muscles affected by poliomyelitis than in normal muscles.

3.4. *The E.M.G. of Elbow Flexors and Extensors*

For the preliminary investigation, normal subjects were used. The subject was seated comfortably with one arm supported with the upper arm fixed in line with the shoulder, so that flexion and extension could be executed in a horizontal plane about a vertical axis through the elbow. In some experiments known loads could be applied at right angles to the wrist, and in others a strain gauge was fixed to the wrist support so that the tension developed during a graded voluntary effort could be measured. A potentiometer on a spindle provided an electrical signal proportional to the angle of flexion.

The forearm support carried a lamp shining a spot of light (the patient's spot) on to a semi-circular scale marked in degrees of flexion set up a few feet away, with the centre of the scale coinciding with the axis of rotation of the elbow. A second lamp was arranged to throw a larger spot of light on the scale (the guide spot) and this guide spot could be moved by an electric motor at a pre-determined angular speed. In experiments at different speeds the subject was required to track the guide spot with his own spot.

The electrodes were silver discs 1 cm in diameter attached by adhesive tape to the skin, 3 cm apart. The subject's skin was prepared by gently rubbing with fine sandpaper and electrode jelly until the resistance was less than 5000 ohms. Two channels were employed. The amplifiers[17] were transistorized, had an input impedance of 100 kΩ, and incorporated a variable bandpass filter, the 6 dB points on the response curves being 20, 60, 150 c/s at the low end and 15 kc/s, 800 c/s and 450 c/s at the high end. The noise level for the band 20 to 800 c/s was 0·25 μV (r.m.s.) with a 5 kΩ wire wound resistor across the input. The e.m.g. was displayed directly on a monitor c.r.o., and was fed via a rectifying and averaging circuit (100 ms time-constant) to a photographic galvanometer recorder. A servo system was used to drive the recording paper. In some cases the input signal to this servo was derived from the potentiometer so that the abscissae represented the angle of flexion. In other experiments it was obtained from the strain gauge so that the abscissæ represented the developed force.

Figure 4 is a recording which shows the relationship between the e.m.g. and the angle of flexion, for a constant load of 4 kg applied at right angles to the wrist, corresponding to a moment of 100 kg cm. When the upper arm muscles are relaxed, the load pulls the arm into a slightly hyperextended position, so that the graph commences on the left of the arm-straight position. As the biceps muscle is contracted, it assists the passive elastic forces in the tissues and the forearm begins its movement. The biceps e.m.g. then increases and so presumably does the tension in the muscle, as the contribution of the passive elastic forces decreases, until a peak is reached at about 10 or 20 deg of flexion. The e.m.g. then begins to decrease until at 80–90 deg it reaches a minimum. It then increases again and eventually, during the last few degrees of flexion, the tissues of the forearm and upper arm are being compressed. Apart from

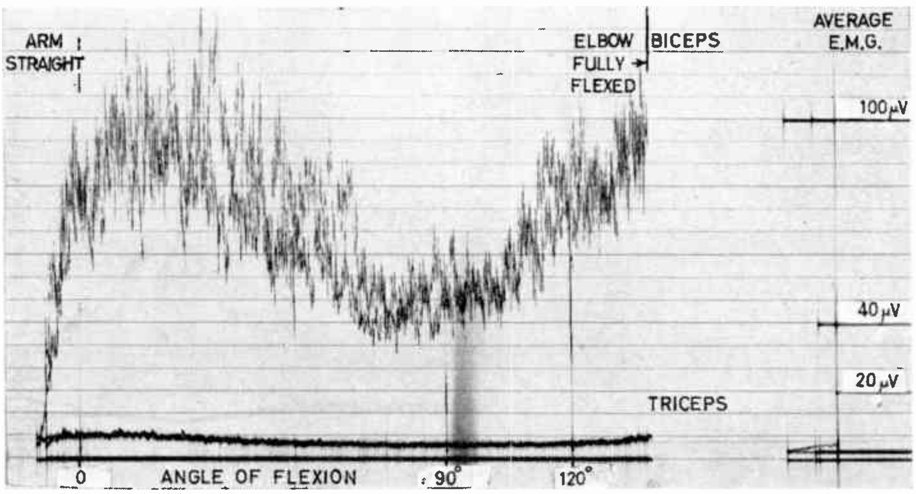


Fig. 4. Recording showing variation of e.m.g. with angle during flexion against a constant load applied at right angles to the wrist. Time for full range of movement approximately 20 seconds. Ordinate: microvolts (r.m.s.). Abscissa: angle of flexion. Two runs superimposed. The lever arm of the load about the elbow was 25 cm. Subject H. Amplifier band 20–800 c/s. Flexing force to wrist was 4 kg.

the effects of these passive elastic forces another factor which contributes to the variation in e.m.g. is the change in the effective leverage of the muscle; the tendon lies a little closer to the axis of the elbow joint when the arm is fully extended or flexed than when it is held at 90 deg. An analysis of this factor has been made by the Los Angeles group.[18] Finally, there is the well-known physiological finding that the tension developed by a muscle, which is artificially stimulated at constant excitation, varies with its length, having a peak near the resting length position and falling off as the muscle is shortened or lengthened from this position.[19] The corollary is that to develop a certain tension the excitation (i.e. e.m.g.) will be least at the resting length position.

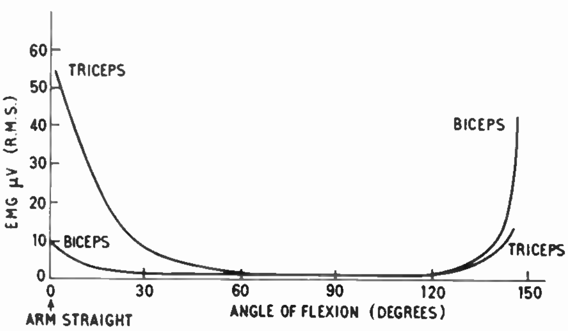


Fig. 5. Tracing taken from a recording for zero load, illustrating effects of tissue distortion at the extremes of the movement, almost complete relaxation of both flexor and extensor groups in the central range, and 'crosstalk' from one group to the other. Subject W. Amplifier band 20–800 c/s.

Figure 5 shows a tracing taken from a recording for zero load. At small angles the triceps is active to straighten the arm against elastic forces, and at large angles the biceps activity is required to flex against elastic forces. In the intermediate range both muscles are almost completely relaxed and the signal approaches the level of the electrical 'tissue noise' approximately 1·5 µV (r.m.s.) as reported by Nightingale.[20]

Figures 4 and 5 show apparent simultaneous activity in both flexor and extensor groups. However, it is very unlikely that both muscle groups are really active at the same time under the conditions of these experiments, and we interpret the effect as 'crosstalk' from one muscle to the electrodes placed over the other. The magnitude of the crosstalk varies with the siting of the electrodes and from one subject to another, but it is usually less than 1/5 of the genuine signal from the active muscle.

Figure 6 is a family of tracings from the biceps muscle obtained from a series of recordings of the type shown in Fig. 4. It is seen that, in the central

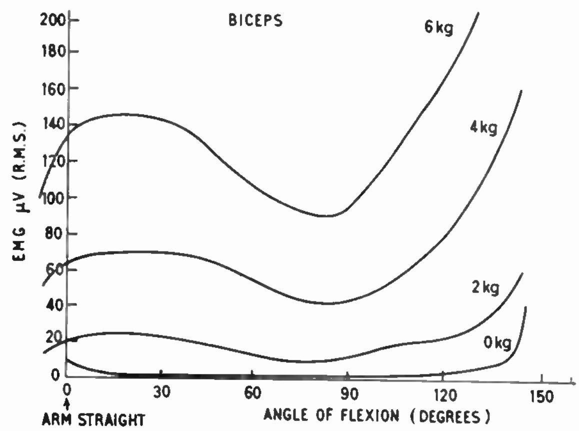


Fig. 6. Family of tracings for the biceps muscle, taken from recordings with various loads applied at right angles to the wrist for a very slow movement. The apparent activity for zero load with the arm straight, is regarded as 'break through' from the extensor muscles, Fig. 5. Subject N. Amplifier band 20–800 c/s.

part of the range of movement, at angles of flexion from about 60 to 100 deg, the e.m.g. goes through a flat minimum.

Figure 7 is a family of curves for the triceps muscle. In this case the movement starts with the arm fully bent, and the effects of tissue compression can be seen at angles near 150 deg. The peaks in the upper three curves and reduction in e.m.g. at 0 deg are attributed to a movement by the subject against his restraining shoulder harness changing the true angle to one in which it was easier to maintain the load. It is seen that the e.m.g. changes relatively slowly with angle over

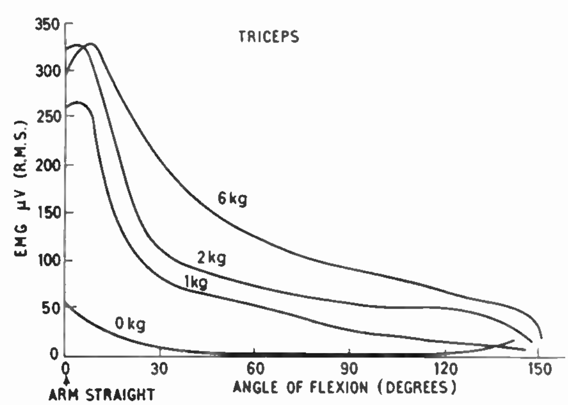


Fig. 7. Family of tracings taken in the same way as those of Fig. 6, for the triceps muscle, for a very slow movement. In this case the movement starts with the arm fully bent. The activity for zero load, with the arm flexed, is regarded as 'break through' from the flexor muscles, cf. Fig. 5. The peaks in the upper three curves near 0 deg are probably due to the tendency of the subject to move slightly against his shoulder-restraining harness. Subject N. Amplifier band 20–800 c/s.

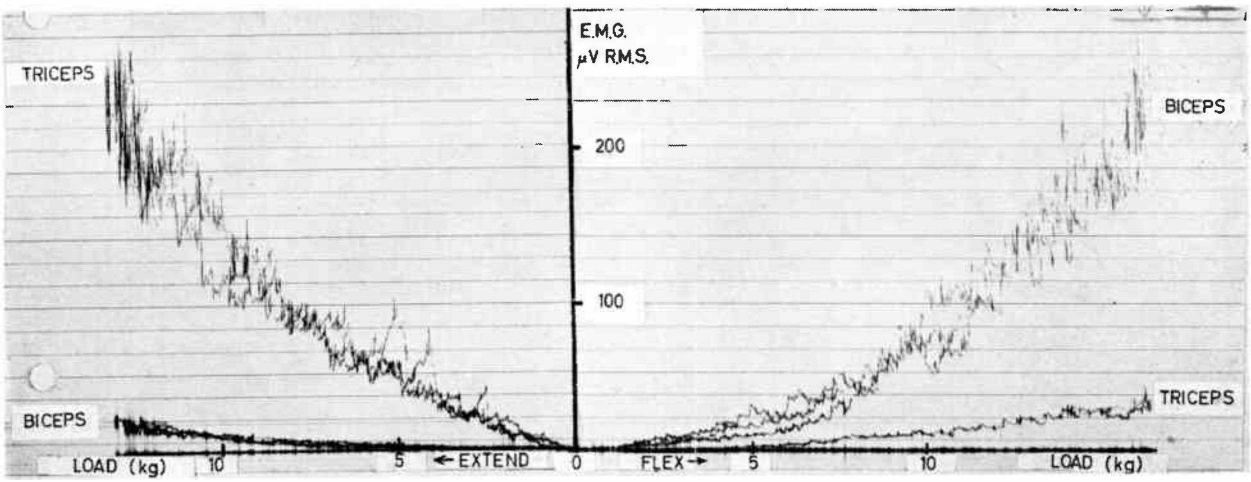


Fig. 8. Relation between e.m.g. and load, applied at right angles to the wrist, under isometric conditions, for gradually increasing effort. Records for 3 runs superimposed. Subject N. Amplifier band 20–800 c/s. Angle of flexion 135 deg.

the range 120 deg to 60 deg. However, there is no minimum as in the case of the biceps.

Figure 8 shows the relation between the e.m.g. and the load under isometric conditions, with the strain gauge 'driving' the recorder. The subject was first asked to make a gradually increasing effort to flex, starting with his muscles relaxed. This drew out the curves on the right, which show the biceps activity with 'crosstalk' to the triceps. The light source of the recording galvanometers was switched off before he relaxed, so that the 'let down' phase of the effort was not recorded. The subject was then asked to make a gradually increasing effort to extend, so that the triceps became active. This is shown to the left of the central line in Fig. 8, with some 'cross-talk' to biceps.

Figure 8 illustrates our general finding that the relationship between e.m.g. and tension for upper arm muscles is not linear, but convex to the tension axis. Various relationships (in some cases linear ones) have been found by previous workers, and have been summarized by Nightingale.[21] From the point of view of using the e.m.g. as a control signal, the non-linearity is not a great disadvantage, provided that there is a clear increase of e.m.g. with tension at all parts of the curve. The tendency, shown in Fig. 8 of the biceps curve to become asymptotic to the load axis near the origin, for this particular subject, is a disadvantage, and led to the investigation of alternative muscles of the flexor group. A better curve was in fact obtained for the brachioradialis muscle. Further investigation of the optimum site for the electrodes is required.

Figure 9 shows the effect of direction of movement. It is seen that the e.m.g. is less during 'let down' when the movement is in the same direction as the load. (This is in general agreement with the results of Bigland and Lippold.[22]) This 'looping' of the e.m.g. angle curve has been found to be almost independent of speed up to 24 deg/s. This is surprising, since it was expected that this effect would be much more marked at higher speeds. Experiments on muscle preparations at constant excitation show a very drastic reduction in the developed force as the speed of shortening is increased (e.g. Hill[23]). Also Bigland and Lippold[22] have reported a considerable increase of the e.m.g. of human muscle with speed of shortening, but very little dependence of the e.m.g. on speed of lengthening. Preparations are in hand to extend these investigations to higher speeds.

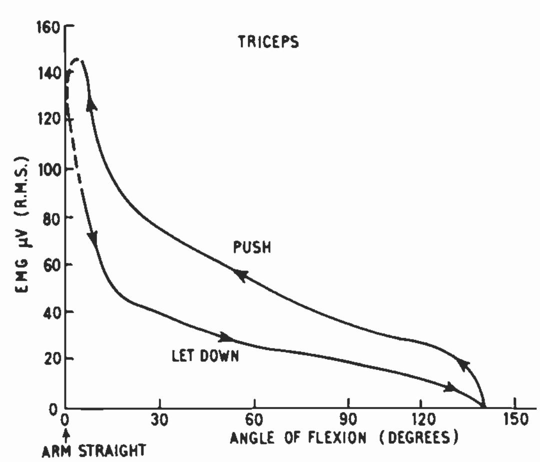


Fig. 9. Showing 'hysteresis' effect; the e.m.g. for a given load is less when the movement is in the same direction as the load. Load 1–5 kg. applied at right angles to the wrist. Amplifier band 20–800 c/s. Constant speed 6 deg/s.

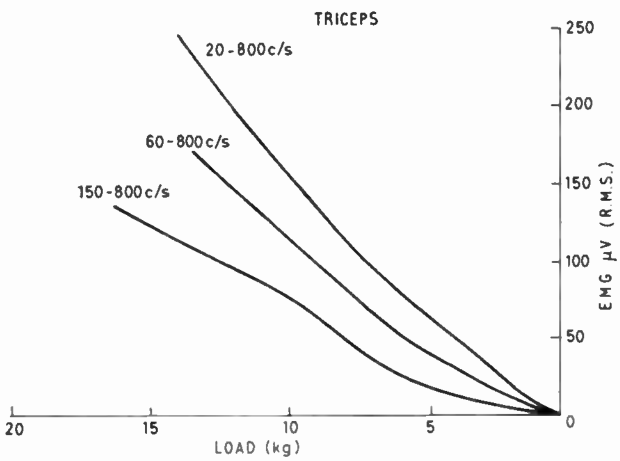


Fig. 10. Showing the effect of restricting the bandwidth of the amplifier at the lower end. Curves taken during gradually increasing effort. (The e.m.g. during decreasing effort was slightly less.)

Preliminary experiments on the effect of the bandwidth of the amplifier showed that extending the high frequency response above 800 c/s simply added to the amplifier noise without increasing significantly the signal amplitude. Figure 10 shows the results of experiments in which the low frequency response was restricted. The reduction in signal amplitude may be worth while accepting under conditions of mains interference, and it is envisaged that a practical amplifier to be used by a patient may have a steeply falling characteristic below 100 c/s. This is in agreement with results previously obtained by Battye, Nightingale and Whillis,[11] and with the e.m.g. spectra reported by Nightingale,[20] and Hayes.[24]

3.5. *Fluctuations and Choice of Smoothing Time-constant*

The e.m.g. waveform shows large irregular fluctuations in amplitude (Fig. 3). Even after rectifying and smoothing (100 ms time-constant) considerable fluctuations remain (e.g. Figs. 4 and 8). The fluctuations could of course be further smoothed out by using a longer time-constant, but this would result in a sluggish response when the e.m.g. is used in a control system. The time-constant of 100 ms was chosen as a reasonable compromise. However, it is undesirable that the fluctuations in the e.m.g. control signal should be reproduced as a trembling in the actuator output, and a method of avoiding this by introducing a degree of backlash is described in the next section.

It is an interesting fact that the fluctuations are not reproduced mechanically in a normal muscle. This is partly because the mechanical twitch of each muscle fibre is of much longer duration than its action potential, and also because the mechanical effects of all active fibres are summated at the tendon, so that there is no 'sampling error' as in the e.m.g. The mechanical inertia and damping of the moving parts will of course also contribute a smoothing effect.

4. Model System

The working model to be described was constructed with the idea of trying to deal with only a few of the problems involved in the use of e.m.g. from surface electrodes in a control system. The task was made artificially easy, firstly by arranging that both the subject and the artificial arm should work over a small change of angle, against a tension which increased rapidly with displacement, secondly a normal subject was used, and thirdly no attempt was made at miniaturization.

Figure 11 shows how the two arms were arranged. The subject was provided with a handle from which springs stretched upwards and downward to fixed moorings. Thus he was able to produce graded tension in either the biceps or triceps group. The motor was provided with a similar task as it followed the movements of the subject's arm.

In this model tension was assumed to be proportional to the integrated e.m.g.

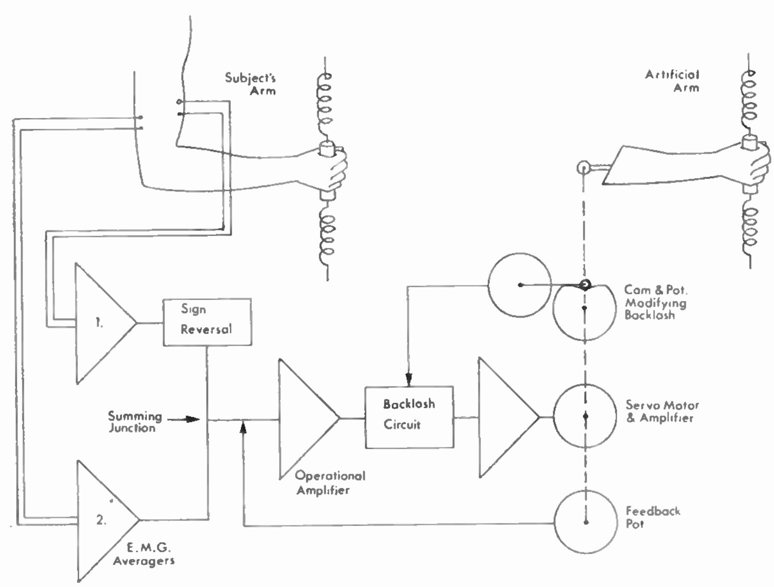


Fig. 11. The model system.

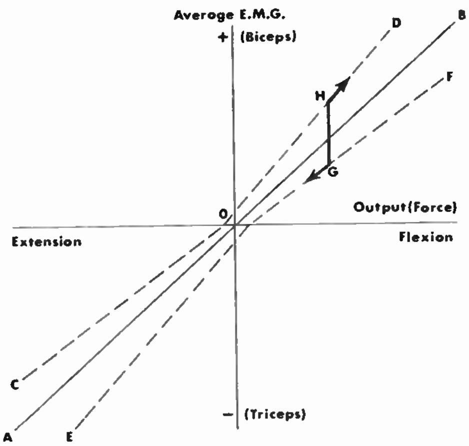


Fig. 12. Illustrating the characteristics of the control system, including the effect of backlash.

and other variables such as angle and velocity were ignored, as was the non-linearity of the relationship. Two difficulties remain and are best explained by reference to Fig. 8. Firstly, there is the problem of the 'crosstalk' already discussed in Section 3.4 above.

Since only one of the opposing muscles is active at a time and the other is always apparently producing what is really crosstalk and hence a lower voltage, we may obtain an estimate of real activity by using the difference in voltage between them. Since we also require to distinguish between the directions of attempted pull and push it is convenient to reverse the sign of one of the channels and add it to the other, giving a voltage proportional to the difference and related in sign to the muscle group whose voltage predominates.

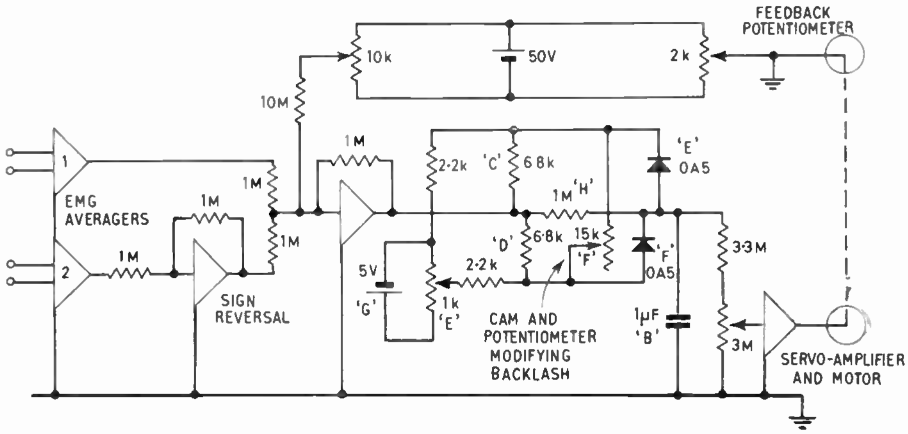
The original idea was to use the resulting voltage as the error signal for a servo-motor giving a characteristic for the system like the line AB in Fig. 12. It was found necessary to complete the servo-loop however because frictional losses in the gears between the motor and the arm prevented the arm returning to the position of zero pull. Ideally, the necessary feedback voltage should have been derived from a tension transducer in the artificial arm, but for simplicity's sake a potentiometer was used (the 'feedback pot' shown in Fig. 11), which was also driven by the motor and hence indicated displacement. This was considered reasonable because tension is proportional to displacement in the arrangement used.

The second difficulty in using e.m.g. from surface electrodes is the variability of the integrated e.m.g. during steady mechanical effort by the muscle, discussed in Section 3.5 above. As can be seen in Fig. 4 this increases as the force rises.

Now if the linear characteristic AB in Fig. 12 were followed these random variations would produce a matched wobble in the motor output. One easy way to reduce this effect would be to increase the time-constant of the averager; on the other hand this would mean a slow response to larger changes in the e.m.g. corresponding to real changes in muscle tension.

Instead of this, a circuit was used having a dual characteristic, the 'backlash circuit' in Fig. 11. The line AB in Fig. 12 is followed but with a slow time-constant, about 1 second. The second characteristic forms an envelope around the first as shown by the broken lines CD and EF and has an almost instantaneous time-constant. This means that short term variations, between these two limiting lines, such as between H and G produce no change in output, whereas if larger short term changes occur the output follows the lines CD and EF as shown. The limits of this envelope are decided by our experience of the amplitude of random variations found experimentally, so that variations greater than this, even if they are sudden may be assumed to be due to a real change in muscle tension. The 'cam and pot modifying backlash' in Fig. 11 is a mechanical method of widening the distance between the lines CD and EF

Fig. 13. Circuit diagram of the backlash circuit.



as force increases to allow for the increasing random variation of e.m.g. with force shown in Fig. 8.

Figure 13 shows the circuit of the complete equipment. The input to the servo amplifier is a function of the voltage across the capacitor B which for small changes in signal is charged and discharged through resistor A. Signal changes in excess of the back-bias voltages set up across diodes E and F charge or discharge capacitor B through resistors C and D giving a faster time-constant. The back-bias (back lash) level is set initially by potentiometer E and raised by force-actuated potentiometer F.

5. Conclusions

The paper has described some of the work which has been carried out in the field of power driven artificial arms. Whilst this work is of a preliminary nature the authors are of the opinion that devices of this kind show considerable possibilities in assisting disabled persons. Even though myo-electric signals are difficult to interpret they can be employed to provide proportional graded control in practice.

6. Acknowledgments

Thanks are due to the following: C. K. Battye, R. E. George and P. Cross, of the Physics Laboratory, Guy's Hospital Medical School, who designed and built the frequency selective pre-amplifiers; P. Styles, of the Medical Electronics Laboratory, St. Thomas' Hospital, for the 'backlash circuit'; and R. Halls, of the Medical Unit, St. Thomas' Hospital, for help and advice.

7. References

1. "The Application of External Power in Prosthetics and Orthotics". Report on the Conference held at Lake Arrowhead California, September 1960. Publication 874, Printing & Publishing Office, National Academy of Sciences—National Research Council, Washington 25, D.C.
2. S. W. Alderson, "The electric arm", Chapter 13 in "Human Limbs and their Substitutes", Edited by P. E. Klopsteg and P. D. Wilson. (McGraw-Hill, New York, 1954.)
3. A. E. Kobrinskii, S. V. Bolkhoivin, L. M. Voskoboinikova, D. M. Ioffe, E. P. Polyan, Ya. L. Slavictskii, A. Ya. Sysin and Ya. S. Yakobson, "Problems of bioelectric control". Proceedings of the International Federation on Automatic Control Conference, Moscow, 1960, pp. 1119–22. (Butterworth, London, 1961.)
4. E. Marquardt and O. Hafner, "The Pneumatic Arm". World Veterans Federation Rehabilitation Bulletin No. 15, pp. 11–24. Also *Archiv. für orthopaedische und Unfall—Chirurgie*, **48**, p. 115. (In German, translation available from the International Society for the Welfare of Cripples, 701, 1st Avenue, New York 17, New York, U.S.A.)
5. S. Weil, "The Heidelberg pneumatic arm prosthesis", *Med. Klin.* No. 7, pp. 242–5, 1959. (In German.)
6. R. Snelson and J. Conry, "Recent advances in functional arm bracing correlated with orthopaedic surgery for the severely paralysed upper extremity", *Orthopaedic & Prosthetic Appliance J.*, **12**, pp. 41–8, 1958.
7. L. A. Geddes, A. C. Moore, W. A. Spencer and H. E. Hoff, "Electropneumatic control of the McKibben synthetic muscle", *Orthopaedic & Prosthetic Appliance J.*, **13**, pp. 33–6, 1959.
8. E. A. Kiessling, "Pneumatic prosthetic components—rigid servo mechanisms and their functional control valves", in reference 1, pp. 116–31.
9. A. B. Kinnier Wilson, "The Hendon pneumatic prosthesis motor", *Brit. Med. J.*, 16th July 1960, p. 227.
10. N. Berger and C. R. Huppert, "The use of electrical and mechanical muscular forces for the control of an electrical prosthesis", *Amer. J. Occupational Therapy*, 6, No. 3, pp. 1–5, May-June 1952.
11. C. K. Battye, A. Nightingale and J. Whillis, "The use of myo-electric currents in the operation of prostheses", *J. Bone Jt. Surg.*, **37B**, pp. 506–10, 1955.
12. A. Weltman and J. Lyman, "The Effects of Electronic Transformations on the Pattern of Myoelectric Activity during Arm Movement". Biotechnology Laboratory Report No. 6, University of California, Los Angeles, September 1960.
13. J. B. Reswick, "Evaluation of digitally programmed control in ortheses and prostheses". In Progress Report No. 1 Medical Engineering, Case Institute of Technology, Cleveland, 6, Ohio.
14. F. Buchthal, "The general concept of the motor unit". In "Neuromuscular Disorders", Vol. 34, pp. 3–30, 1961. Research Publication of the Association for Research in Nervous and Mental Disease.
15. F. Erminio, F. Buchthal and P. Rosenfalck, "Motor unit territory and muscle fibre concentration in paresis due to peripheral nerve injury and anterior horn cell involvement", *Neurology*, 9, No. 10, pp. 657–71, 1959.
16. J. A. R. Lenman, "Quantitative electromyographic changes associated with muscular weakness", *J. Neurol. Neurosurg. Psychiat.*, **22**, pp. 306–10, 1959.
17. C. K. Battye, "A variable band-width transistor preamplifier for electromyography". To be published.
18. G. E. Gwynne, "Kinesiology of elbow flexion". In Biotechnology Laboratory Progress Report, Dec. 1960. University of California, Los Angeles.
19. H. J. Ralston, "Uses and limitations of electromyography in the quantitative study of skeletal muscle function", *Amer. J. Orthodontics*, **47**, pp. 521–30, 1961.
20. A. Nightingale, "Background noise in electromyography" *Phys. Med. Biol.*, **3**, pp. 325–38, 1959.
21. A. Nightingale, "The graphic representation of movement. II. Relationship between muscle force and the electromyogram in the stand-at-ease position", *Ann. Phys. Med.*, **5**, pp. 187–91, 1960.
22. B. Bigland and O. C. J. Lippold, "The relation between force, velocity and integrated electrical activity in human muscles", *J. Physiol.*, **123**, pp. 214–24, 1951.
23. A. V. Hill, "The mechanics of voluntary muscle", *Lancet*, **1951 ii**, pp. 947–51, 24th November 1951.
24. K. J. Hayes, "Wave analysis of tissue noise and muscle action potentials", *J. Appl. Physiol.*, **15**, pp. 749–52, 1960.

Manuscript first received by the Institution on 27th June 1963, and in final form on 27th September 1963. (Paper No. 86/MBE11.)

U.D.C. 621.372.45: 621.372.542

The Effects of Error in the Element Values and the Converter Transfer Characteristic on the Responses of Active RC Filters

By

A. G. J. HOLT, Ph.D., Dip.El.
(*Associate Member*)†

AND

F. W. STEPHENSON, B.Sc.†

Summary: This paper first deals briefly with negative impedance converters and with the methods of active network synthesis devised by Linvill and Yanagisawa. The greater part of the paper treats the effects of small errors in the element values and in the transfer characteristic of the negative impedance converter on the response of low-pass Maximally Flat Magnitude (Butterworth) and Equal Ripple (Chebyshev) filters of second-order. Butterworth and Chebyshev filters of the fourth-order are also considered. Curves are given showing the way in which the errors distort the response characteristics; the importance is pointed out of the proximity to the negative impedance converter of a circuit element having an error in its value. A symmetry between the effects of error in circuit elements on either side of the converter is also noted. The method of display of the sensitivity of circuits to errors used in this paper is believed to be unique for active networks in that the effect on the response curve is shown directly. In other work on the subject, reviewed in this paper, equations are given showing the sensitivity of the poles of the network function to changes in element values and converter sensitivity. The pole shifts are, of course, a measure of the effect on the response of the filter but the effect is shown more clearly when a change in the response curve is given.

1. Introduction

One important topic within the domain of electrical network theory which has received increasing attention in recent years is the synthesis of filter networks using resistors, capacitors and active elements. These networks can be designed to meet prescribed excitation-response characteristics similar to those which can be obtained from the more familiar circuits employing inductors, capacitors and resistors. The advantages of eliminating inductors from the filters are particularly apparent at frequencies of less than about 50 c/s. At these frequencies suitable inductors become bulky and expensive, are likely to be of comparatively low Q factor and pick up hum.

The use of resistors and capacitors in filters is attractive because they are smaller, cheaper and more nearly ideal elements than are inductors. When employed in conjunction with transistors these components can give highly selective, compact filters for use at low frequencies. The need for a power supply for the transistors is unimportant if, as frequently happens, the filter is part of a large apparatus for which a power pack is required regardless of the needs of the filter.

The work of Linvill,[1] Yanagisawa,[2] Kinariwala,[3] Sandberg[4] and Sipress[5] in the U.S.A. has enabled active filters to be designed using a negative impedance converter (n.i.c.) as the active element. A n.i.c. is a two-port (two terminals at input, two at output) which offers an input impedance of $-Z_L$ at one port when an impedance Z_L is connected to the other port. Several methods have been obtained whereby active networks giving required response characteristics may be synthesized; that due to Yanagisawa is perhaps the simplest to use, while the technique developed by Linvill is the basic method from which others have evolved. A brief outline of these synthesis methods will first be given.

2. Types of Negative Impedance Converter

Two types of negative impedance converter exist, both of which offer an impedance $-Z_L$ at the input port when the other port is loaded with an impedance Z_L. For one variety known as the voltage inversion type and shown in Fig. 1, the current at the output port is equal to the current at the input port while the

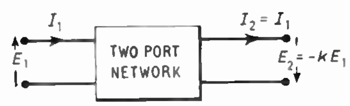


Fig. 1. Voltage inversion type of negative impedance converter.

voltage at the output port is related to the voltage at the input port by a constant $-k$. Hence the voltage is inverted through the converter. The equations relating the voltages and currents for the voltage converter are

$$E_1 = -kE_2 + 0.I_2, \quad I_1 = 0.E_2 + 1.I_2 \qquad \begin{bmatrix} E_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} -k & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_2 \\ I_2 \end{bmatrix} \quad \ldots\ldots(1)$$

† Electrical Engineering Department, University of Newcastle upon Tyne.

The other variety of converter, known as the current inversion type, is shown in Fig. 2. For this the voltage at the output port is equal to that at the input port while the current at the output port is related to the current at the input port by a constant $-k$. Hence the current is inverted through the converter. The equations relating the voltages and currents for the current converter are

$$\begin{aligned} E_1 &= 1.E_2+0.I_2 \\ I_1 &= 0.E_2-kI_2 \end{aligned}, \quad \begin{bmatrix} E_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -k \end{bmatrix} \begin{bmatrix} E_2 \\ I_2 \end{bmatrix} \qquad \ldots\ldots(2)$$

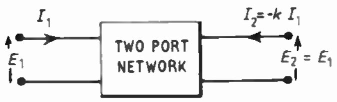


Fig. 2. Current inversion type of negative impedance converter.

The converters shown in Figs. 1 and 2 are assumed to be ideal but in the practical case it is found that small departures from the ideal occur. Larky [6] has discussed the effect of these departures on the action of the converter and also given a method whereby compensation may be applied by the connection of additional series and shunt elements to the converter.

3. Yanagisawa's Method of Synthesis

This method[2] is also known as the parallel *RC* active network synthesis. It uses a negative impedance converter of the current inversion type and two *RC* networks to give a voltage transfer ratio E_1/E_2; the general form of the final network is shown in Fig. 3 where the converter is assumed to have a k value of $+1$. Frequently the two passive *RC* circuits are taken to be in the form of inverse L networks as shown in Fig. 4. When this is done the synthesis of a required voltage transfer ratio can be carried out as follows.

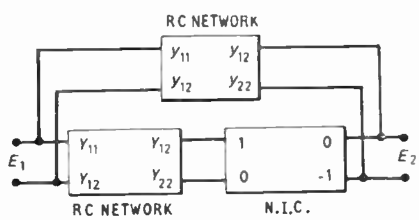


Fig. 3. Basic Yanagisawa circuit.

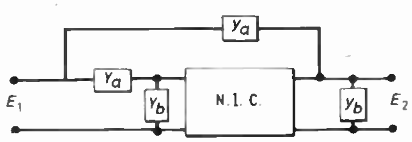


Fig. 4. Yanagisawa circuit using L networks.

For Fig. 3 the voltage transfer ratio can be written as the ratio of two polynomials, thus

$$\frac{E_1}{E_2} = \frac{N(p)}{D(p)} \qquad \ldots\ldots(3)$$

This can also be written

$$\frac{E_1}{E_2} = \frac{D(p)+N(p)-D(p)}{D(p)} \qquad \ldots\ldots(4)$$

For the circuit employing two L networks the voltage transfer ratio is

$$\frac{E_1}{E_2} = \frac{y_a-Y_a+y_b-Y_b}{y_a-Y_a} \qquad \ldots\ldots(5)$$

It is necessary to select $(n-1)$ points σ_i on the negative real axis of the p plane, where n is the highest order of the numerator or denominator polynomials in E_1/E_2.

Let the arbitrary polynomial be formed

$$K(p) = \prod_{i=1}^{n-1}(p-\sigma_i) \qquad \ldots\ldots(6)$$

Dividing both numerator and denominator of eqn. (4) by $K(p)$ gives

$$\frac{\dfrac{D(p)+N(p)-D(p)}{K(p)}}{\dfrac{D(p)}{K(p)}} \qquad \ldots\ldots(7)$$

Comparing terms between eqn. (5) and eqn. (7) gives the admittances

$$y_a-Y_a = \frac{D(p)}{K(p)} \qquad \ldots\ldots(8)$$

$$y_b-Y_b = \frac{N(p)-D(p)}{K(p)} \qquad \ldots\ldots(9)$$

The L networks may be formed from eqns. (8) and (9) by letting the terms with positive residues appear in y_a and y_b while those with negative residues appear in Y_a and Y_b.

4. Linvill's Method

This technique is also known as the positive *RC* negative *RC* synthesis. It yields a network giving a required transfer impedance $Z_{21} = E_2/I_1$ using one negative impedance converter of the voltage inversion type and two passive *RC* circuits. The general arrangement is shown in Fig. 5 where the converter is again assumed to have a k value of -1.

Analysis of the network in Fig. 5 shows that the transfer impedance is

$$\left.\frac{E_2}{I_1}\right|_{I_2=0} = Z_{21} = \frac{Z_{12a}.Z_{12b}}{Z_{22a}-kZ_{11b}} = \frac{N(p)}{D(p)} \qquad \ldots\ldots(10)$$

One can carry out the synthesis by selecting an arbitrary polynomial $K(p)$, as was done for the Yanagisawa method, and dividing the numerator and denominator polynomials of Z_{21} by $K(p)$.

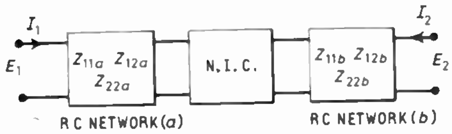


Fig. 5. Basic Linvill circuit.

Here $K(p) = \prod_{i=1}^{n} (p-\sigma_i)$ where n is not less than the degree of $N(p)$ or $D(p)$, whichever is greater.

If the expression for $D(p)/K(p)$ is expanded it is found to be of the form

$$\frac{D(p)}{K(p)} = \sum \frac{A_i}{(p-\sigma_i)} \qquad \ldots\ldots(11)$$

The terms with positive residues may be realized by Z_{22a} and those with negative residues by Z_{11b}. Factors of $K(p)$ can be allocated between Z_{12a} and Z_{12b} when the required distribution of positive and negative terms between Z_{22a} and Z_{11b} has been carried out. Once the driving point impedances of the (*a*) and (*b*) networks are known, the passive *RC* circuits can be synthesized using, for example, the Cauer method to give simple *RC* ladders. This is the method which was employed to design the network having the Maximally Flat Magnitude (Butterworth) and Equal Ripple (Chebyshev) characteristics described in later sections of this paper.

5. Previous Work on the Sensitivity of Networks

It is evident that when active devices are included in filters the possibility arises that their characteristics will drift slowly with time or as a result of temperature changes. This is an important problem which has received some attention in the literature. In his basic paper on active filters Linvill[1] considered the effects of slow drifts of the converter performance on the natural frequencies of the network, pointing out that a change in these natural frequencies can be translated into a change in the filter response. Linvill gives an expression for the pole shift in a natural frequency in terms of the impedance change ΔZ and the derivative with respect to p of the denominator impedance in eqn. (10).

In a paper on negative impedance converters Larky[6] deals with the effects of parameter drift on the input impedance of a converter which has been compensated to behave as an ideal component. Larky draws conclusions showing which parasitic elements in the converter circuit should be made small in order to reduce the effects of parameter drift.

Blecher[7] and Sipress[5] give expressions for the change in the position of a transfer function pole produced by a change in a circuit element. They give the result that unless the magnitude and direction of the pole movement can be selected independently, attention should be concentrated on minimizing the magnitude of the pole movement without regard to the direction. They also give a table of approximate optimum pole sensitivities for second order functions.

The synthesis methods outlined in Sections 3 and 4 both require the selection of an arbitrary polynomial and it is evident that the exact polynomial chosen must have a considerable effect on the resulting network design. Horowitz,[8] in his paper on optimization, shows how polynomials may be selected which give the minimum sensitivity to changes in the element values and to drifts in the transfer characteristic k of the converter. The polynomials employed in the filter designs given in later sections of this paper were chosen with reference to Horowitz's work.

None of the papers quoted above show the effect of changes in element values or of changes in the transfer characteristic of the converter on the response of typical filters. All the results are in the form of equations and no curves are drawn showing to what extent a filter response alters when, for example, the value of k changes by 10%. In this paper designs are given for second-order and fourth-order Butterworth (maximally flat magnitude) and Chebyshev (equal ripple) filters. The effects of changes in circuit elements and in the converter transfer characteristic on the response curves are shown in the form of distorted response curves and as error curves.

6. Filter Designs

6.1. *Second-Order Functions*

Two relatively simple second-order functions were synthesized and produced networks of the same general form (Fig. 6), from which it can be shown that

$$\frac{E_2}{I_1} = Z_{21}$$

$$= \frac{R_3}{p^2C_1C_2R_2R_3+p[C_1R_2+R_3(C_2-kC_1)]+1} \qquad \ldots\ldots(12)$$

where p is the complex frequency variable. In each of the second-order responses considered, the correct curve is obtained when $k = 1$. The effect of variations

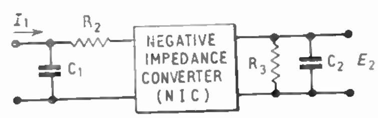


Fig. 6. Simple second-order function synthesized network.

of k, above and below unity, on the frequency characteristic ($|Z_{21}|_{p=j\omega}$) and the effect of element variation, with $k = 1$, were investigated with the aid of a digital computer.

As far as the shape of the characteristic is concerned, variation of R_2 has the same effect as variation of C_2 whilst variation of R_3 corresponds to changing C_1. In addition, alteration of the value of R_3 alters the level of the curve (flat loss).

6.1.1. Butterworth response

The transfer impedance

$$Z_{21} = \frac{H}{p^2+1{\cdot}414214p+1} \qquad \ldots\ldots(13)$$

where H is a constant multiplier, yielded the network shown in Fig. 7.

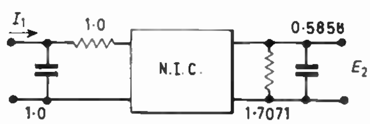


Fig. 7. Butterworth second-order filter.

All element values are normalized with respect to unity cut-off frequency and unit resistance. (See Appendix.)

A sample of the results from this network is shown in Figs. 8 (*a*), (*b*) and (*c*).

6.1.2. Chebyshev response

The transfer impedance (for $\varepsilon^2 = 0{\cdot}995262$)

$$Z_{21} = \frac{H}{p^2+0{\cdot}6449p+0{\cdot}707948} \qquad \ldots\ldots(14)$$

yielded the network shown in Fig. 9 (again all element values are normalized).

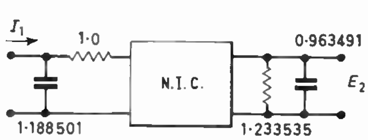


Fig. 9. Chebyshev second-order filter.

A sample of the results from this network are shown in Figs. 10 (*a*), (*b*) and (*c*).

6.2. *Fourth-Order Functions*

6.2.1. Butterworth response

A fourth-order function of the Butterworth form was synthesized using a single n.i.c., the resulting filter

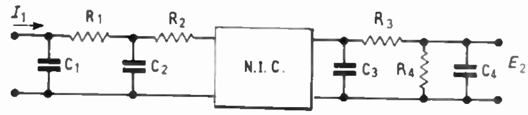


Fig. 11. Butterworth fourth-order function synthesized network.

being as shown in Fig. 11. Analysis of this network yields the following expression for the transfer impedance:

$$Z_{21} = \frac{H}{Ap^4+Bp^3+Cp^2+Dp+1} \qquad \ldots\ldots(15)$$

where

$$H = R_4, \qquad A = C_1C_2C_3C_4R_1R_2R_3R_4$$

$$\begin{aligned}
B = \{&C_3C_4R_3R_4[R_2(C_1+C_2)+C_1R_1]+\\
&+(R_4C_3+R_4C_4+R_3C_3)C_1C_2R_1R_2-\\
&-kC_1C_2C_4R_1R_3R_4\}\\
C = \{&[R_4(C_3+C_4)+C_3R_3][C_1(R_1+R_2)+C_2R_2]+\\
&+C_1C_2R_1R_2+C_3C_4R_3R_4-\\
&-k[C_1C_2R_1(R_3+R_4)+\\
&+R_3R_4C_4(C_1+C_2)]\}\\
D = \{&[R_4(C_3+C_4)+C_3R_3]+\\
&+[R_2(C_1+C_2)+C_1R_1]-\\
&-k(C_1+C_2)(R_3+R_4)\}
\end{aligned}$$

As in the second-order examples, the desired response is obtained when $k = 1$ and the effect of variations was investigated by means of a digital computer.

As far as the shape of the curve is concerned,

(i) variation of R_1 is equivalent to variation of C_4
(ii) ,, ,, R_2 ,, ,, ,, ,, ,, C_3
(iii) ,, ,, R_3 ,, ,, ,, ,, ,, C_2
(iv) ,, ,, R_4 ,, ,, ,, ,, ,, C_1

In addition, alteration of the value of R_4 will change the flat loss.

The transfer impedance for the Butterworth fourth-order function is

$$Z_{21} = \frac{H}{p^4+2{\cdot}613126p^3+3{\cdot}414214p^2+2{\cdot}613126p+1} \qquad \ldots\ldots(16)$$

which yielded the network shown in Fig. 12. A sample of the results obtained from this network is shown in Figs. 13 (*a*)–(*e*).

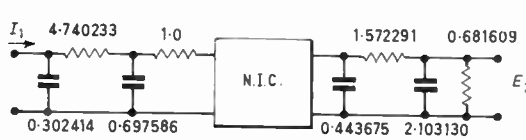


Fig. 12. Butterworth fourth-order filter.

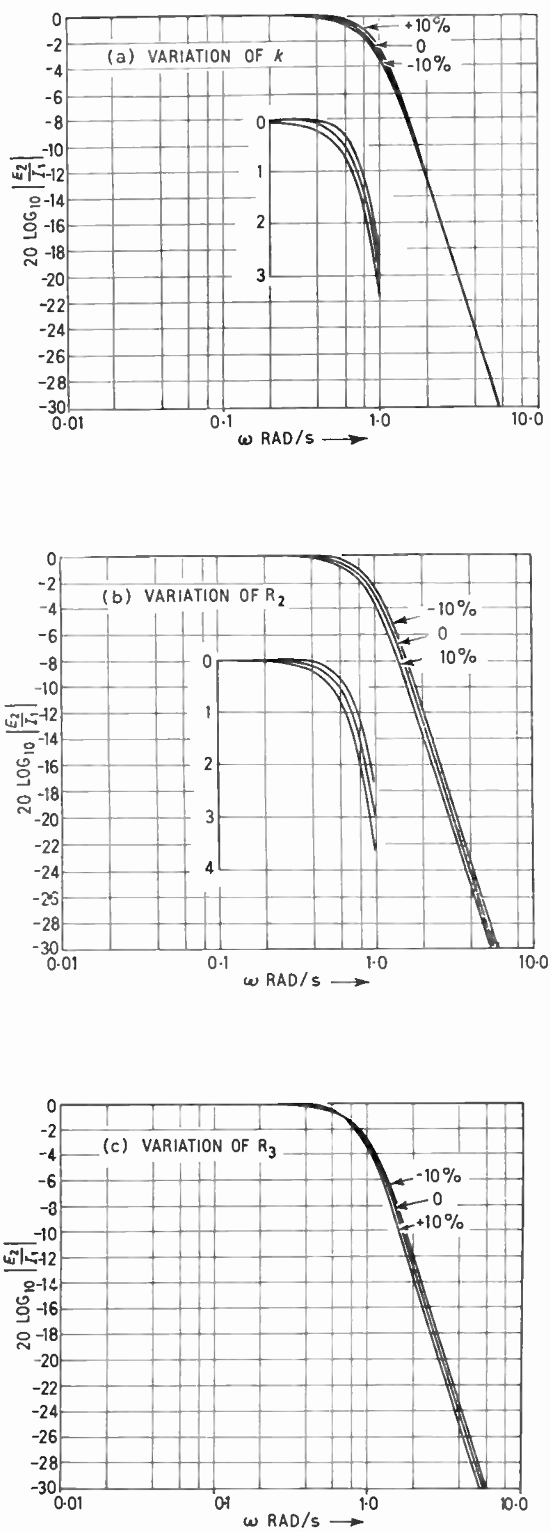


Fig. 8. Butterworth second-order filter. Insets show curves in the region $\omega < 1{\cdot}0$ to a larger scale.

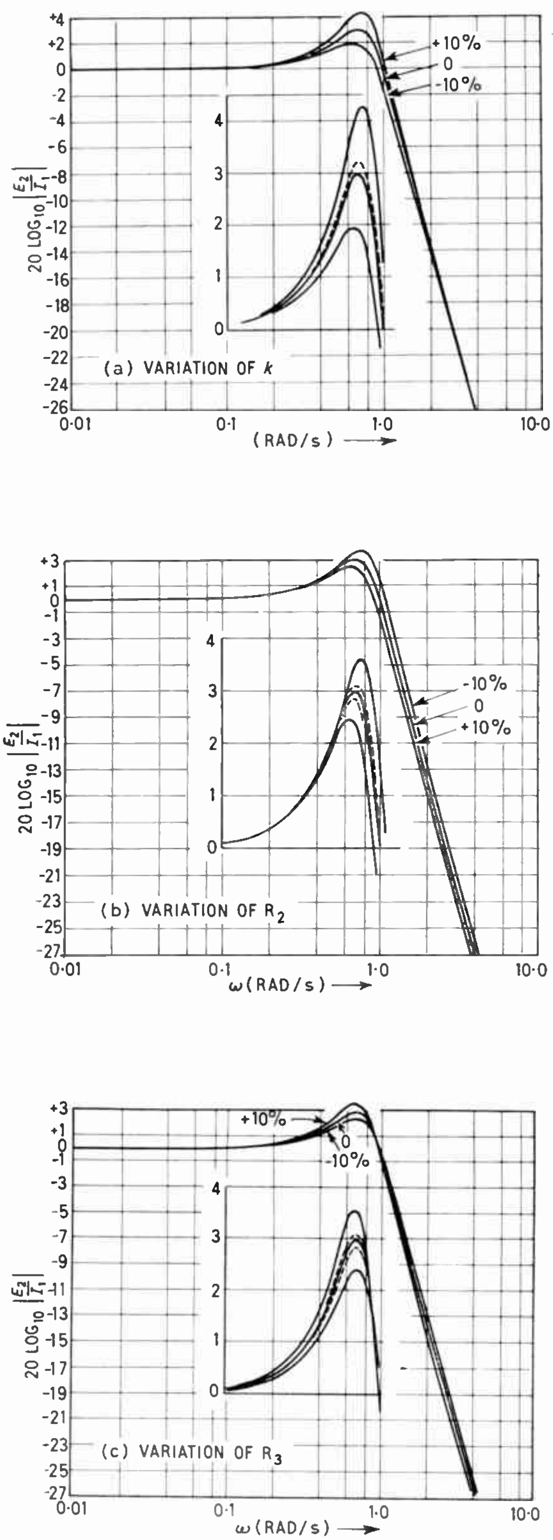


Fig. 10. Chebyshev second-order filter. Insets show curves in the region $\omega < 1{\cdot}0$ to a larger scale. Broken curves represent 2% variation.

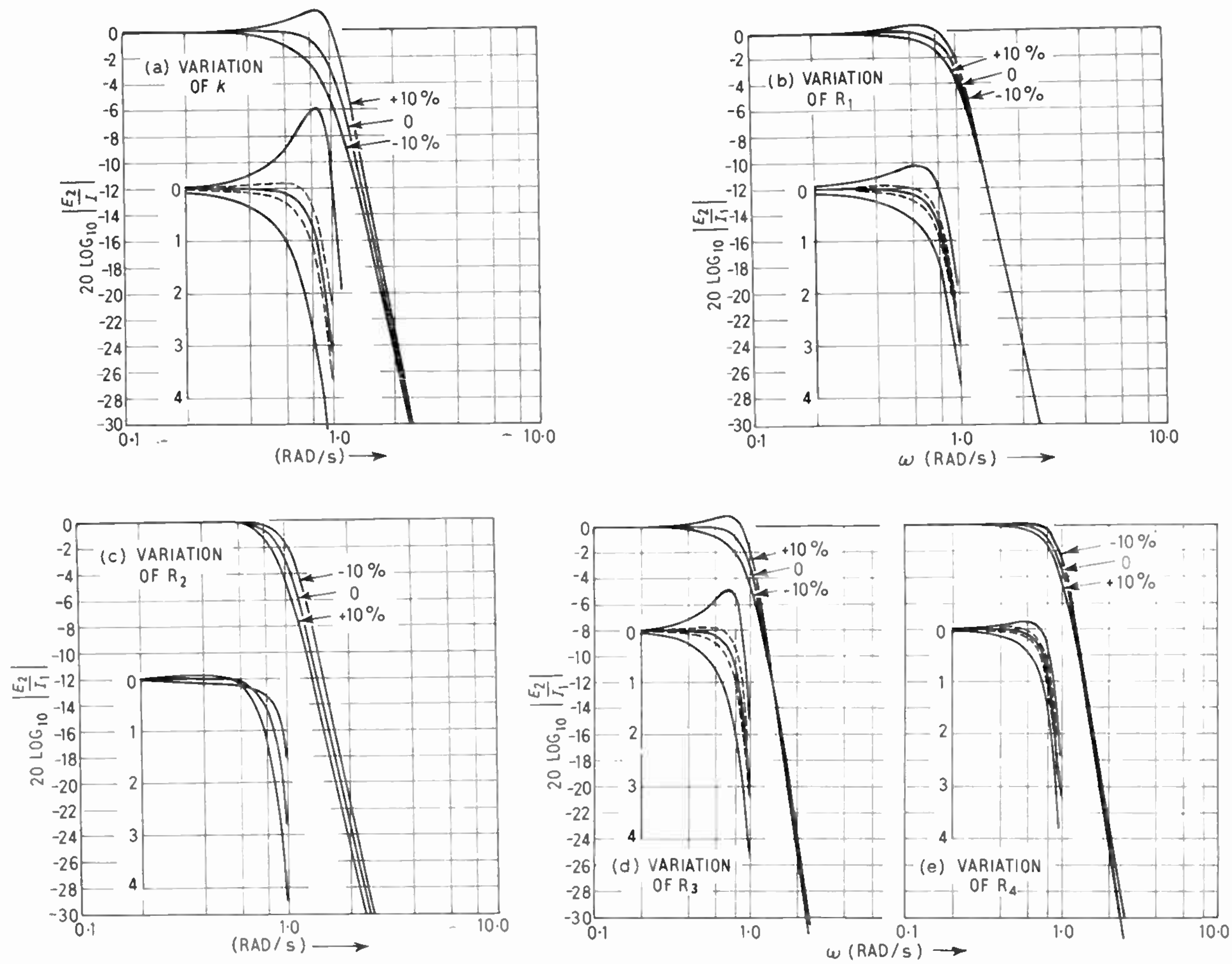


Fig. 13. Butterworth fourth-order filter. Insets show curves in region $\omega < 1{\cdot}0$ to a larger scale. Broken curves represent 2% variation.

6.2.2. Chebyshev response

A procedure similar to that for the fourth-order Butterworth filter was followed with a fourth-order Chebyshev design having a pass-band ripple amplitude of $20 \log_{10} \left|\frac{E_2}{I_1}\right| = 3$ dB. The transfer impedance function is

$$Z_{21} = \frac{H}{p^4 + 0{\cdot}581580p^3 + 1{\cdot}169118p^2 + 0{\cdot}404768p + 0{\cdot}176987} \qquad \ldots\ldots(17)$$

The form of the network is identical to that of the Butterworth filter shown in Fig. 11. Element values for the Chebyshev design are:

$R_1 = 4{\cdot}582795$	$C_1 = 0{\cdot}872248$
$R_2 = 1$	$C_2 = 0{\cdot}594648$
$R_3 = 2{\cdot}388102$	$C_3 = 0{\cdot}29005$
$R_4 = 3{\cdot}502941$	$C_4 = 1{\cdot}141136$

Resistances are in ohms and capacitances in farads.

Small changes were made in these element values and in the converter transfer characteristic k, the responses were computed and a selection from the results is shown in Figs. 14 (*a*)–(*e*).

7. Derived Results

From the results (a selection of which are illustrated in Figs. 8, 10, 13 and 14), it is possible to investigate the absolute error $\left(\text{in units of } 20 \log_{10} \left|\frac{E_2}{I_1}\right|\right)$ as a function of ω. Curves are shown in Figs. 15 and 16 for the second-order functions.

Similarly, Figs. 17 and 18 illustrate the errors in the Butterworth and Chebyshev fourth-order filters, respectively. Figure 19 compares the sensitivity to conversion factor for the cases of second- and fourth-order Butterworth characteristics and Fig. 20 gives comparable curves for the Chebyshev characteristics. The examples are all for 10% variation in element or

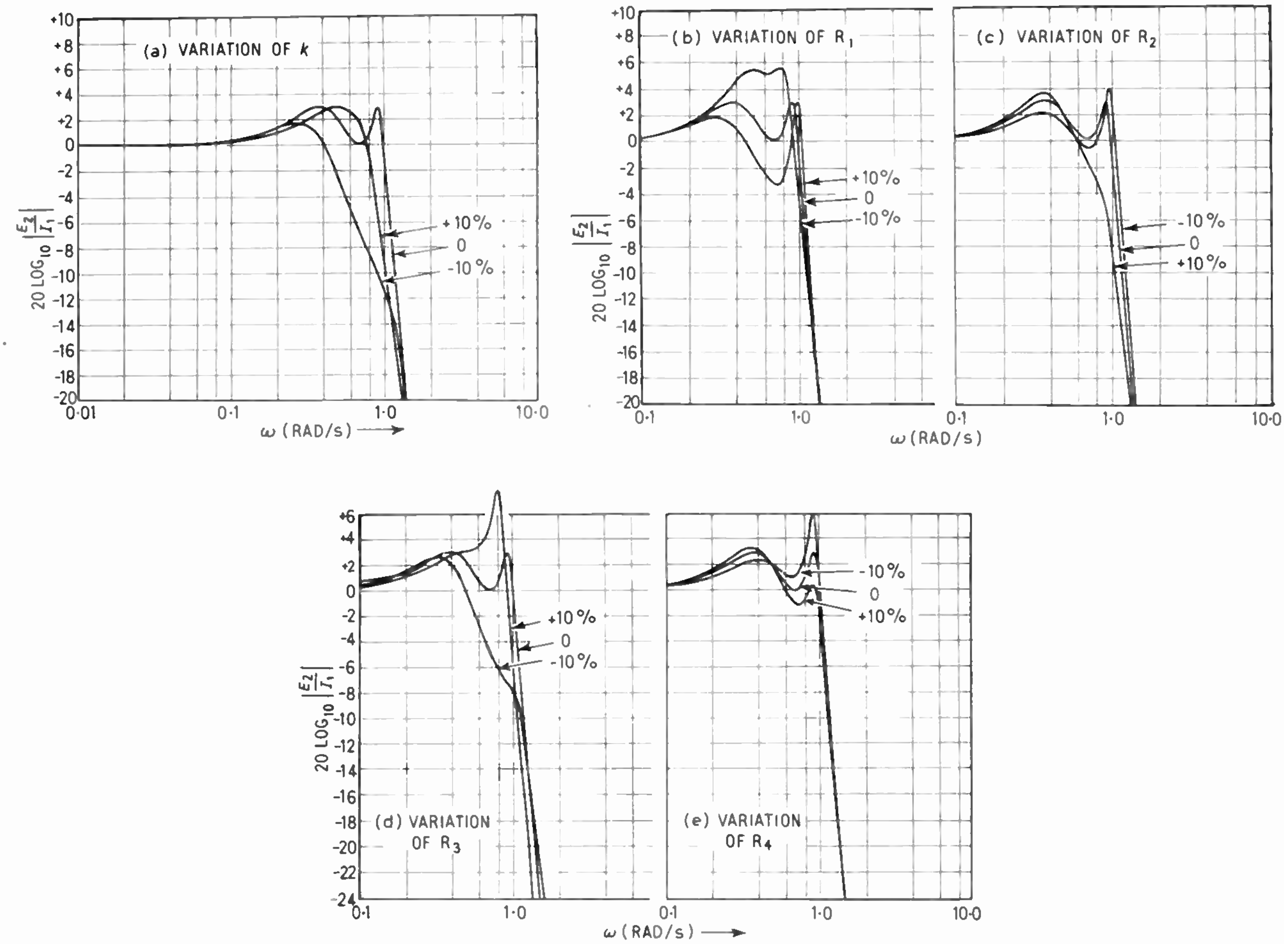


Fig. 14. Chebyshev fourth-order filter.

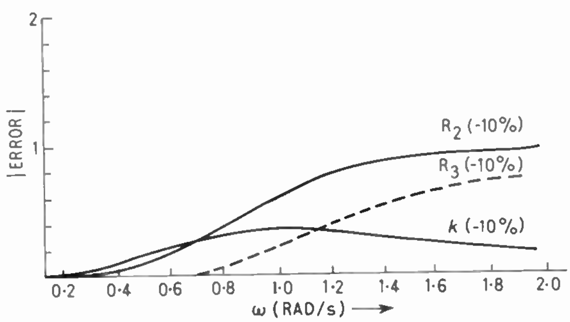


Fig. 15. Butterworth second-order filter. Error $\left(\text{in units of } 20 \log_{10} \left|\frac{E_2}{I_1}\right|\right)$ against ω for n.i.c. and filter elements.

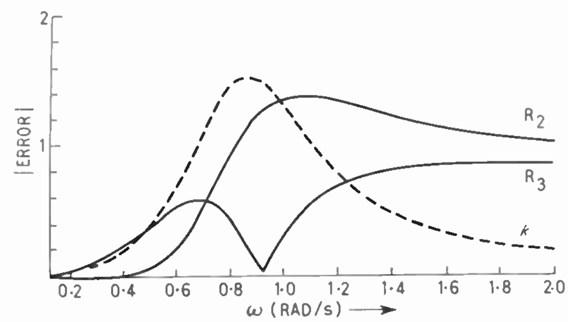


Fig. 16. Chebyshev second-order filter. Error against ω for n.i.c. and filter elements.

conversion factor which, although larger than one would expect in practice, serve the purpose of providing greater curve separation.

8. Conclusions and Further Work

A very important problem which arises in the practical design of active filters has been shown to be that of sensitivity to active and passive element changes. Using the Horowitz method, the sensitivity can be minimized for a given function when one n.i.c. is used. From the curve in Fig. 16, it is well illustrated that the response is much more dependent upon k in the case of a fourth-order function than that of a second-order. In terms of the error units used, the

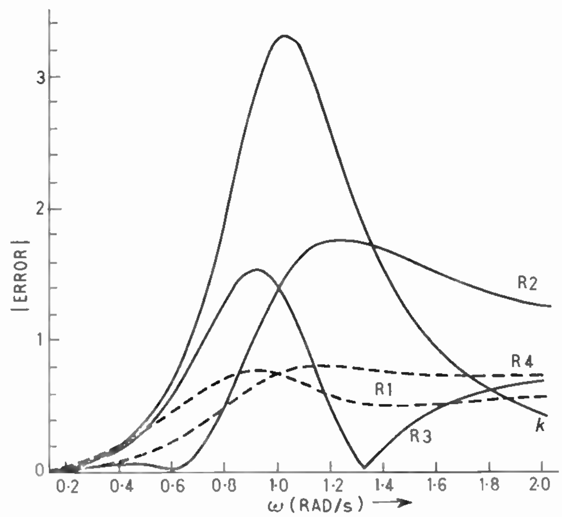


Fig. 17. Butterworth fourth-order filter. Error against ω for 10% variation in n.i.c. and filter element.

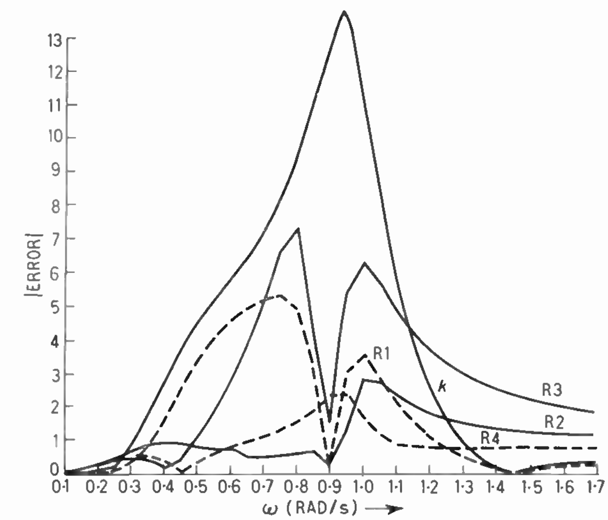


Fig. 18. Chebyshev fourth-order filter. Error against ω for 10% variation in n.i.c. and filter elements.

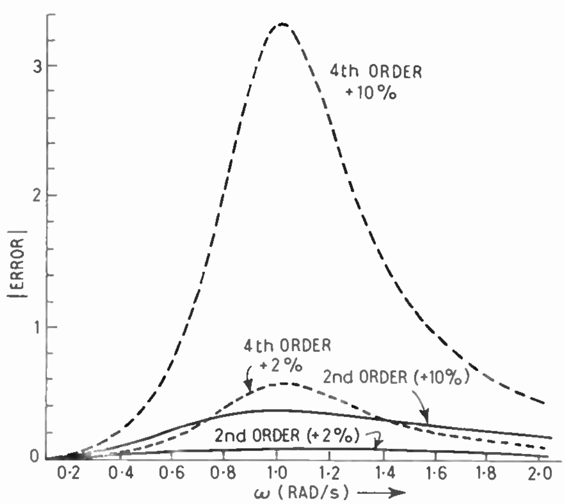


Fig. 19. Derived results showing error produced by variation of converter transfer characteristic in second- and fourth-order Butterworth filters.

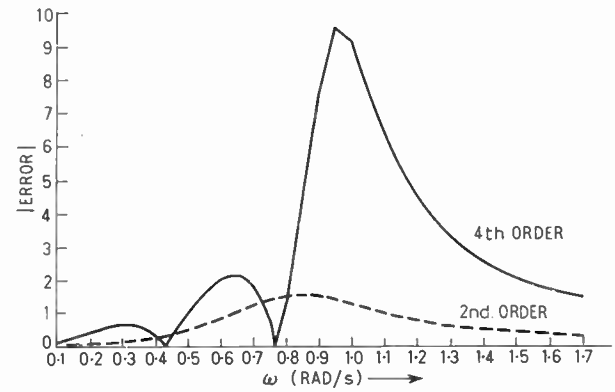


Fig. 20. Derived results showing error produced by variation of converter transfer characteristic in second- and fourth-order Chebyshev filters.

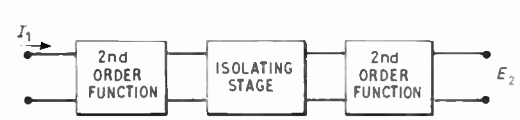


Fig. 21. Fourth-order function synthesized as cascade of two second-order stages.

difference in the error, for the 10% and 2% curves at $\omega = 1$, is approximately 8 : 1.

If, however, a fourth-order function were to be synthesized as a cascade of two second-order stages (Fig. 21) then the overall sensitivity of the network would be the sum of the individual errors. In the examples shown this would mean a decrease in sensitivity of the order of 1 : 4 as compared with a stage using one n.i.c.

Figures 17 and 18 show the error sources to be grouped into three parts:

(i) The n.i.c., by far the most sensitive

(ii) R_2, R_3 forming the mid-group

(iii) R_1, R_4 forming the least sensitive group.

It would seem from this that the effect of elements upon the characteristic is dependent in some way upon their proximity to the n.i.c.

No conclusions can be drawn as to the generality of the results obtained at this stage although it should be noted that practical work, which initiated this investigation, has suggested that the sensitivities in the two methods for producing fourth-order functions would differ quite markedly.

Work is at present being carried out on other synthesis methods in order to see whether a generalized pattern can be found along the broad lines illustrated in this paper.

9. Acknowledgments

The authors have pleasure in thanking Professor R. L. Russell, Professor of Electrical Engineering in the University of Newcastle upon Tyne, for his interest and for the use of the facilities of the Electrical Engineering Department. Thanks are also due to the staff of the University Computing Laboratory for their help.

One author (F. W. S.) has pleasure in acknowledging the financial support of a research studentship awarded by the Welwyn Electric Co. Ltd., Bedlington, Northumberland.

10. References

1. J. G. Linvill, "RC active filters", *Proc. Inst. Radio Engrs*, **42**, pp. 555–64, March 1954.
2. T. Yanagisawa, "RC active networks using current inversion types negative impedance converters", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-4**, pp. 140–4, September 1957.
3. B. K. Kinariwala, "Synthesis of active RC networks", *Bell Syst. Tech. J.*, **38**, pp. 1269–1316, September 1959.
4. I. W. Sandberg, "Synthesis of transformerless active *N*-port networks," *Bell Syst. Tech. J.*, **40**, pp. 761–83, May 1961.
5. J. M. Sipress, "Synthesis of Active RC Networks". D. E. E. Dissertation, Polytechnic Institute of Brooklyn, New York, June 1960.
6. A. I. Larky, "Negative impedance converters", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-4**, pp. 124–31, September 1957.
7. F. H. Blecher, "Application of synthesis techniques to electronic circuit design", *Intl Conv. Rec. Inst. Radio Engrs*, **8**, Part 2, pp. 210–222, 1960.
8. I. M. Horowitz, "Optimization of negative-impedance conversion methods of active RC synthesis", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-6**, pp. 296–303, September 1959.

11. Appendix: Denormalization

It should be noted that the element values in filters shown in Figs. 7, 9 and 12 are normalized with respect to unit resistances and unity cut-off frequency. In the actual filter it may be convenient to give R_2 a value of $10^3\Omega$. This would be accomplished by multiplying each impedance element in the filter by the resistance denormalization factor, 10^3. Since the network elements are linear, this change does not alter the character of the response and amounts to a scaling of the magnitude function by 10^3. Since resistances are directly proportional to impedances, they are multiplied by 10^3 whilst capacitance values are divided by 10^3 (being inversely proportional to impedance).

If it be required that the cut-off point ($\omega = 1$) be at 10^3 c/s (for instance) then this can be accomplished by a frequency denormalization $2\pi \times 10^3$. To keep the impedances of all elements invariant with the change of variable as the shift from normalized frequency to denormalized frequency occurs, all capacitances are divided by the denormalization factor.

Manuscript received by the Institution on 29th July, 1963. (Paper No. 867.)

New Aerial Tuning System for G.P.O. Rugby V.L.F. Transmitter

Automatic tuning and phase correction has now replaced the manually-controlled variometer in the 400-kW very-low-frequency transmitter (GBR) at the G.P.O. Rugby Radio Station. Rugby has been in service for nearly 40 years and continues to carry mainly morse telegraph traffic but also transmits precise time signals four times each day. The 16-kc/s carrier frequency is derived from a 100-kc/s Essen-ring quartz-crystal oscillator, which is stable to within a few parts in 10^{10} each day. These signals are observed by many organizations concerned with the measurement of time, and in particular the carrier is used for the inter-comparison of national frequency standards.†

High winds, humidity and temperature variations and especially icing and low clouds, influence the electrical characteristics of the extensive aerial system, which is supported on 12 masts, each 820 ft in height and spaced at quarter-mile intervals. These weather effects cause de-tuning of the aerial circuit which, if left uncorrected, results in changes of up to $\pm$ 45 deg in phase of the radiated carrier accompanied by a reduction in output power up to 50%. In the past, a manually-controlled variometer, forming part of the aerial tuning inductance in the transmitter, has been adjusted by an operator to keep the aerial in tune and maintain maximum aerial current. This method is adequate for only fairly gradual long term effects and is in any case difficult to use when the transmitter is being keyed. Rapid fluctuations which can be set up by strong winds disturbing the aerial for instance, are too irregular to be corrected in this way.

These irregular fluctuations in radiated power due to aerial detuning are obviously undesirable since they affect the efficiency of the transmitter in traffic. For those concerned with the precise measurement of frequency, the phase variations are equally objectionable, since they limit the accuracy with which the primary frequency standards can be observed.

The new variometer is a servo-driven type, controlled by a phase-sensitive discriminator which compares the phase of the aerial current against a reference derived from the output of the final amplifier. The d.c. output from the discriminator is applied to a 50-c/s transistor modulator which drives the servo-motor and thereby adjusts the setting of the variometer. Velocity-controlled feedback is applied to the system by a tacho-generator output, to ensure freedom from overshoot.

The variometer has an inductance range of 51 μH (min. 41 μH; max. 92 μH). It is driven by a 28-W a.c. two-phase servomotor and can traverse the full range in 30 seconds. The variometer is wound with 6561-strand 36 s.w.g. 'litz' cable rated at 1000 A, accommodated in cages of s.r.b.p. tubes mounted on end-plates of similar material. The rotor weighs nearly 1 cwt (50 kg) and the final drive is imparted via a notched nylon-loaded rubber belt and toothed s.r.b.p. pulley, as metal cannot be employed in the vicinity of the coils. Special p.t.f.e. bearings are used in the rotor to keep down frictional losses and the precision worm-drive gearing of the motor drive is virtually free from backlash. This is necessary to overcome the motoring torque caused by the aerial current flowing through the variometer.

The automatic aerial tuning system is capable of responding to changes in tune during a single telegraph element of 20 ms duration and has a residual phase error not exceeding $\pm$ 1 deg. Also, the main aerial tuning coils, which are made of 6561-strand 'litz' cable wound on wooden spiders approximately 16 ft in diameter, have been stiffened to minimize a 'phase jitter' of up to 30 deg caused by electro-mechanical vibration of the coils at about 3–5 c/s, during 'on–off' keying. In addition, equipment has been installed which compares the phase of the aerial current relative to the crystal oscillator and corrects any slow overall changes in phase, such as changes in early stages of the transmitter.

The combined effect of these innovations is to maintain the amplitude of the aerial current to 0·1% of its maximum value and to keep the phase of the radiated carrier generally within 0·5 deg relative to that of the crystal oscillator with occasional excursions to $\pm$ 1 deg. This will produce a marked improvement in performance of the transmitter, particularly during bad weather, which will be welcomed by the many users of GBR in the scientific field. At the same time the operation of the transmitter will be considerably simplified.

† J. McA. Steele, "Standard frequency transmissions", *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*), **26**, No. 1, p. 78, July 1963.

Post Office Radio Tower

The Post Office's London radio tower, which is being erected alongside the four-storey extension of the Museum Telephone Exchange in Central London, is designed to handle up to 150 000 simultaneous telephone conversations and 40 channels for monochrome or colour-television relay.† Preliminary details of the technical features to be installed have recently been published by the G.P.O.

Existing television radio-relay links to Norwich, Birmingham and Southampton, working from the present 180-ft lattice mast on the old telephone exchange roof, will be transferred to the tower, starting with the Norwich link towards the end of 1964. In 1965 further links will be installed to provide 3600 trunk telephone circuits to Birmingham and to distribute the B.B.C. second television programme to Birmingham, Bristol, Southampton and the Isle of Wight.

The tower has now reached a height of 500 ft (December 1963) and when completed will be 620 ft. high and weigh approximately 13 000 tons.

† See also "Communications developments", *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*), **25**, p. 398, May 1963.

U.D.C. 621.396: 963.325

Recent Trends and Future Developments in Radar Displays

By

D. W. G. BYATT, B.Sc.†

Presented at the Symposium on "Processing and Display of Radar Data" in London on 16th May 1963.

Summary: The paper starts with a brief recapitulation of the development of radar displays, from the simple form of moving coil plan position indicator (p.p.i.) to the development of the d.c.-coupled fixed-coil displays, capable of allowing for target marking during intertrace periods. An extension of this technique uses secondary high-speed deflection coils to provide alpha-numeric identification of the intertrace position.

A description is given of a modern transistorized radar display, capable of performing the dual role of raw marked radar display or synthetic display.

Improvements in cathode-ray tubes, and the need for the bright display is discussed. Particular emphasis is given to the bright direct-view storage tube, and the method of operation and performance of displays using these tubes is given.

Radar to television scan conversion systems are described in both forms, using either the bombardment induced conductivity tube, or the storage vidicon. Some comments are made upon the possible use of new devices such as electroluminescent panels.

1. Introduction

As the narrow beam from a radar aerial rotates, usually relatively slowly, the trace upon the plan position indicator is required to rotate in synchronism upon the cathode-ray tube, so that the radio echoes from the aircraft indicate its position.

Early surveillance radar p.p.i.'s generated the trace upon the c.r.t. by the use of a sawtooth current waveform applied to a deflection coil. Rotation of the trace was achieved by simply rotating the coil on the neck of the tube, in synchronism with the aerial head. A.c. coupling was usually employed, and centrality was maintained by area-balancing the waveform. This type of display is still very commonly used for marine radar.

However, over the past twenty years, the use of radar for tracking aircraft in both the civil and military field has become firmly established, and the need has arisen for marking aircraft on the display by more elegant means than the chinagraph pencil. The most satisfactory method of doing this is to use a fixed coil deflection system, d.c.-coupled throughout. Sawtooth waveforms varying sinusoidally are applied, in quadrature, to the X and Y deflection coils. A rest period between the end of one scan, and the beginning of the next is allowed. This time can be used to produce a mark on the display at a position determined by the X (and Y) d.c. level during this period.

Control of these X and Y levels is obtained from a device such as a joy-stick, attached to potentiometers across a d.c. supply. The level set by these potentiometers is then gated into the deflection waveform prior to the display.

The earlier forms of intertrace marks were simple geometric shapes, obtained by adding suitable deflection waveforms on to the intertrace period. These shapes were most commonly used to indicate particular aircraft to other operators on other displays, but the system was subsequently developed, and additional marks were used to identify aircraft being partially or completely automatically tracked.

The use of this type of intertrace marking has the advantage of very good accuracy. As both the intertrace and sawtooth waveforms traverse the same display system, registration errors between intertrace mark and radar echo can be kept to an order of 0·1% without too much difficulty. Hence it becomes possible to extract the position of an aircraft by an X and Y voltage with respect to an origin, giving in effect, a conversion from polar to cartesian co-ordinates. By suitable encoding, this information can be fed to a digital store and computer, and can form some of the

† The Marconi Company Ltd., Display and Data Handling Laboratories, Great Baddow, Essex.

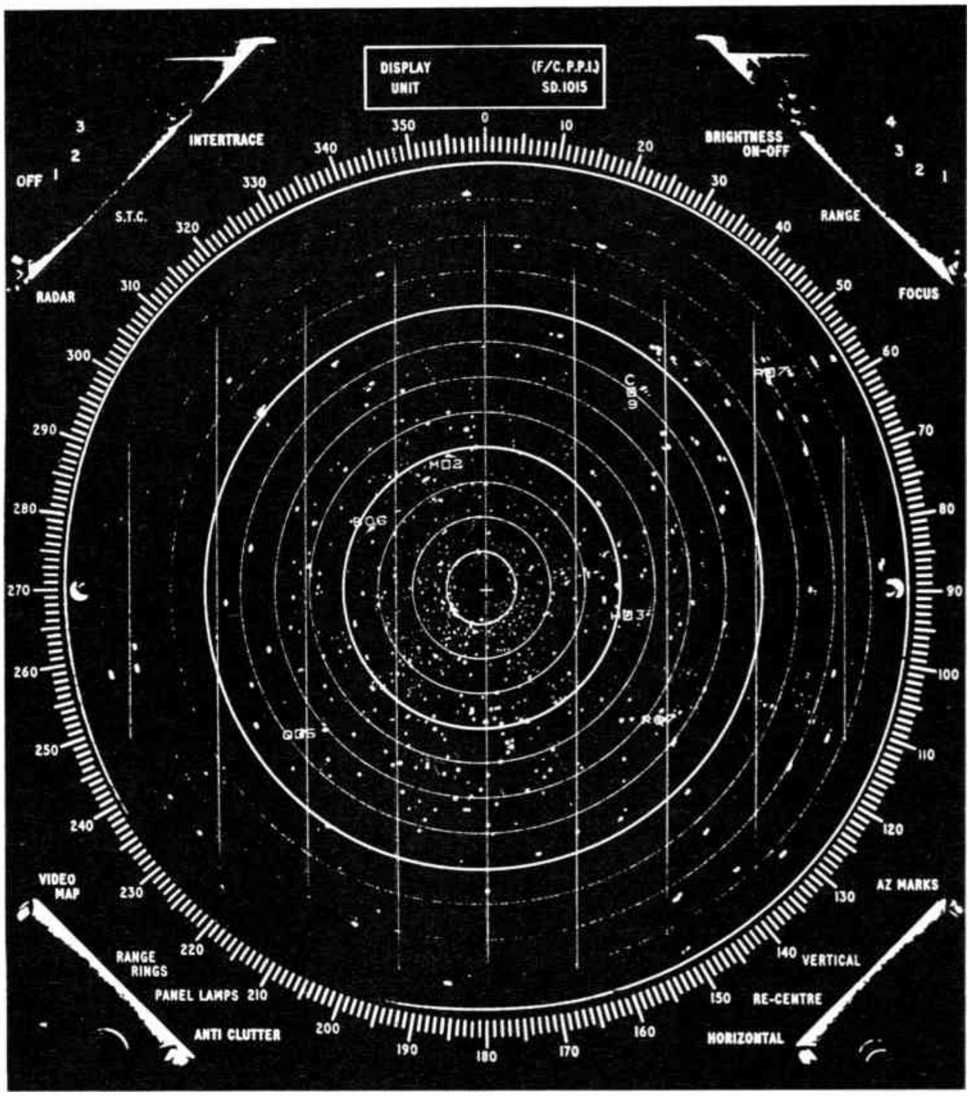


Fig. 1. Labelled raw radar display.

basic data for a computer-aided air traffic control system.

As the numbers of aircraft increased, and data handling systems came into use, it became obvious that more than simple circles or squares were required for marking aircraft, and several methods of writing identification numbers and letters around the aircraft were tried. One of the most satisfactory ways of doing this is to generate small X and Y waveforms which represent the form of letters and numbers, and display these during the intertrace period. However, intertrace time is limited and, in fact, for a given set of conditions, the longer the intertrace period, the less the maximum range of the display. In order to write a sufficient number of characters in one intertrace period, it is necessary to follow small deflections of about 1 μs duration. An example of a marked raw radar display is shown in Fig. 1.

This is best achieved by using the main deflection amplifier to position the c.r.t. beam at the correct d.c. position of the character group, and to use a small aperture high-speed deflection coil behind the main coil, driven by separate amplifiers, for writing the characters. The signal mixing is then done on the c.r.t. beam. By this means it is possible to write a letter or number of extreme clarity in 20 μs.

As previously stated, the position of each aircraft is obtained from the tracking radar operator's display, either by extracting the X and Y co-ordinates of each aircraft by means of the controllable markers, or by automatic tracking, monitored by the operator. This information, together with aircraft identification, height, estimated times of arrivals, etc., is converted to digital form and fed to the computer store for processing. Successive positions of the aircraft enable the velocity to be found. Because of the high operating speed of the computer, negligible delay occurs between the insertion and availability of the data. Once the information is in processed form, it can be made available upon a synthetic display unit in an uncluttered form, with selection carried out as required. Therefore, a comprehensive radar display should be able to accept 'raw' radar, for basic control and/or data extraction, together with the ancillary video information, and also perform the function of a synthetic display for operational decisions.

The essential technical difference between a raw and synthetic display is in the waveforms. The aircraft positions on the synthetic occur on a random staircase waveform, where each level represents the position, in a manner similar to that of intertrace marking.

One of the main objections to the conventional radar display is its inherent lack of brightness. This is mainly because of the low rate of information from the aerial head, which often rotates at only 4 rev/min, but also partly because of the need for afterglow tails. These two requirements necessitate a long afterglow phosphor cathode-ray tube, which can only be viewed satisfactorily under low levels of ambient lighting, and require radar operators to work in controlled lighting conditions, or to use viewing hoods.

For many operational situations, such as in airfield control towers, the bridge of a ship or the cockpit of an aircraft, there is a real need for a bright display. Apart from brightness, the conventional cathode-ray tube provides an ideal radar display. A tube with a usable diameter of 10 in (25·4 cm) can easily provide a spot size of 0·01 in (0·25 mm), picture geometry can be held to adequate limits, and tube costs are relatively low.

2. A Modern Radar Display Unit

The main operational units of a display are:

(1) The main time-base deflection amplifiers.

(2) The high-speed symbol writing amplifiers.

(3) The video amplifier.

Each of these units is complex, and usually performs several functions. There are also ancillary parts, such as e.h.t. generator and regulator, and the focus regulator. An example of the units of a modern radar display is shown in Fig. 2.

2.1. *Time-base Deflection Amplifier*

The main deflection amplifiers of the transistorized display to be described can be separated into two parts. The first sub-amplifier accepts low level floating earth inputs, and provides balanced push-pull outputs. In order that several displays may be operated from one central time-base, and yet show different parts of the radar cover area, it is customary to include off-centring and range expansion, within the display amplifier. This particular amplifier accepts a ± 10 V sawtooth input. This comparatively low voltage has proportionally higher external noise pickup than the previous valve time-base levels and, consequently, a floating earth is carried with the signal, to effect noise cancellation at the long-tail pair input. The amplifier

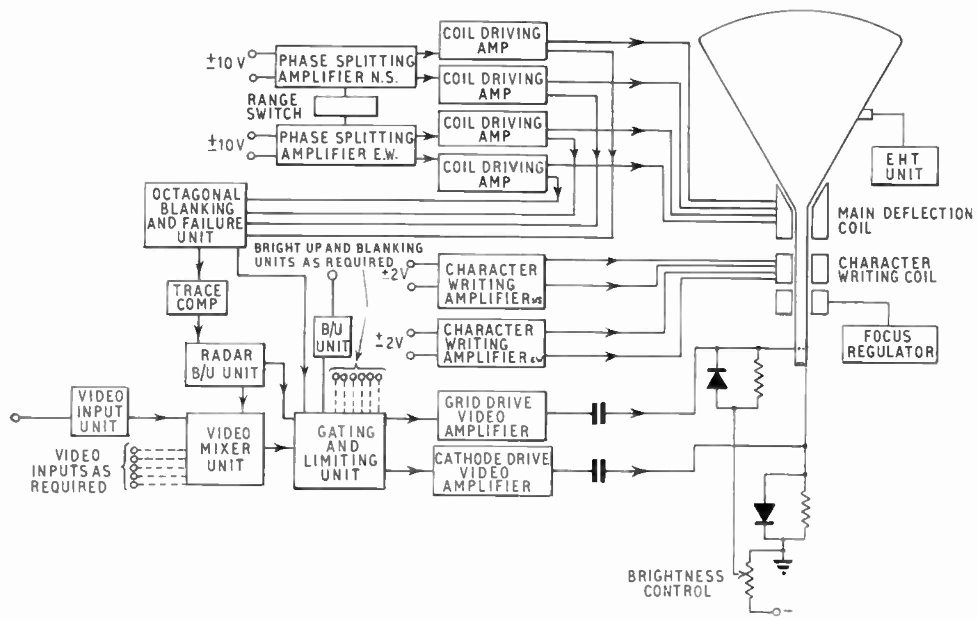


Fig. 2. Transistorized radar/synthetic display.

is in effect, a balanced see-saw amplifier, with three feedback loops, one from each output, and one common mode. Range expansion (of ×1, ×2, ×4, ×8) is obtained by switching the feedback resistance values. Off-centring is achieved by injecting controlled d.c. voltages into another arm of the amplifier. Feedback limiting controls the signal level handled by the amplifier on the higher expansions, and prevents excessive charging of capacitors, which would give sluggish recovery, and excessive heating effect. One of the important considerations in the design of transistorized deflection amplifiers is the effect of temperature. By using this type of balanced amplifier, differential temperature drift is kept to a very small value, the absolute effect being catered for by the use of silicon transistors. In order to minimize out-of-phase cross-coupling effects in the coil, a separate amplifier, with its own feedback loop, drives each half of the deflection coil, as shown in Fig. 3.

Amplifier stability has been considerably eased by splitting the amplifiers into two parts, the first part involving expansion and limiting, but working into a resistive load. The second part has to drive a reactive load, but is a straightforward fixed gain amplifier. The c.r.t. beam can be made to traverse the full tube diameter, and settle to 0·1% of the diameter in 50 μs. The bandwidth of the system from input to c.r.t. face is such that a delay of less than 2 μs occurs. This means a misregistration of intertrace level and time-base range of about 0·15 miles, which can easily be removed by adjustable compensating capacitors in the unit.

2.2. *Symbol Writing Amplifier*

The symbol writing amplifier is more simple in design than the main amplifiers but, as before, uses push-pull circuits with grounded base silicon planar transistors of f_T = 100 Mc/s, for coil driving. These give a small signal bandwidth to the tube face of greater than 2 Mc/s at 3 dB down. With this amplifier, the full deflection amplitudes upon the tube face are so small that drift and distortion become second order and there is little difficulty in achieving excellent results.

2.3. *Video Amplifier*

The video amplifier is divided into three parts. The output unit drives the grid and cathode of the c.r.t. with push-pull signals of 60 V amplitude. Prior to this amplifier is the gating unit, which accepts bright-up signals and blanking signals, for example character bright-up, or video blanking.

To cater for the high symbol writing speeds, this unit, together with the output amplifiers, provides rise-times of about 0·1 μs. To allow for varying operational conditions, up to six bright-up and blanking inputs can be accepted. The third unit of the video amplifier is the video mixing unit. Once more, up to six video inputs are provided, for radar, secondary radar, video map, etc., each with a separate control, and each channel having a bandwidth of 4·5 Mc/s through to the c.r.t.

The video amplifier also has waveforms supplied from the main deflection amplifiers that blank the video beyond the periphery of the c.r.t. face, to prevent flaring, and also scan failure signals that blank the video in the event of failure of time-base signals up to and including the deflection coils. The bunching of the radial scans near the centre of rotation on the display necessitates a compensating bright-up waveform. This, together with a waveform to correct for

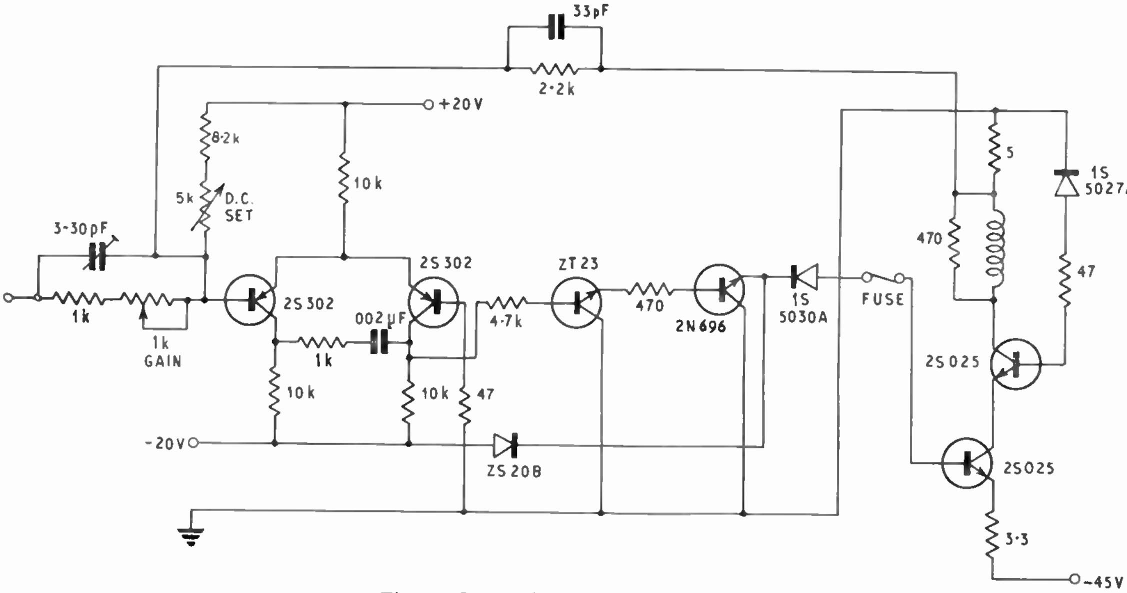


Fig. 3. One main coil driving amplifier.

degrees of expansion, is also incorporated in the video amplifier. Adjustable video limiting circuits enable a range of signal/noise ratios to be painted upon the cathode-ray tube.

The e.h.t. unit operates at ultrasonic frequency to provide 14 kV $\pm$ 0·5% from a self-oscillatory circuit. The advent of high performance semiconductors has made possible the design and production of a flexible radar display unit giving superior performance in time-base speeds and video characteristics than were previously obtainable with practical valve circuits, and with a four-to-one reduction in power consumption and heat dissipation.

3. Bright Radar Displays

3.1. *The Direct-view Storage Tube as a Radar Display*

The cathode-ray tube used in the display previously described is a more highly developed version of a basic tube used for radar over the past two or three decades. Main improvements have been to spot size and phosphor. Light is emitted from the phosphor where it is struck with sufficient velocity by the electron beam. The most intense light is given off in the first flash, thereafter light emission continues, at a much lower level, and falling exponentially in amplitude.

The direct-view storage tube produces a bright display by continually flooding electrons through a storage mesh which has the radar information traced upon it.[1] The essential parts of the English Electric E.702 5-in direct-view storage tube are shown in Fig. 4. The viewing screen is a high-efficiency short-persistence phosphor on the interior of the face plate. The target structure, **B**, consists of an extremely fine metal mesh with a high insulation material deposited upon it on the side nearest the flood gun F. It is the capacitance between the front surface of the insulating material and target mesh that provides the storage property of this tube.

The flood gun electrons are collimated by electrodes to arrive normal to the screen, and to illuminate the whole area uniformly.

The tube electrode potentials are such that the flood gun electrons strike the storage mesh at a velocity at which the secondary emission ratio for the insulator is less than unity. The mesh surface therefore becomes negatively charged to a level close to that of the flood gun cathode. The flood gun electrons penetrate the mesh, and uniformly illuminate the screen. The storage mesh is primed for writing upon by applying an external voltage to take the mesh more negative than the cathode. This causes the flood gun electrons to be repelled, and they are collected on g_4.

The high velocity electrons from the writing gun are focused to a fine spot, and can be modulated and

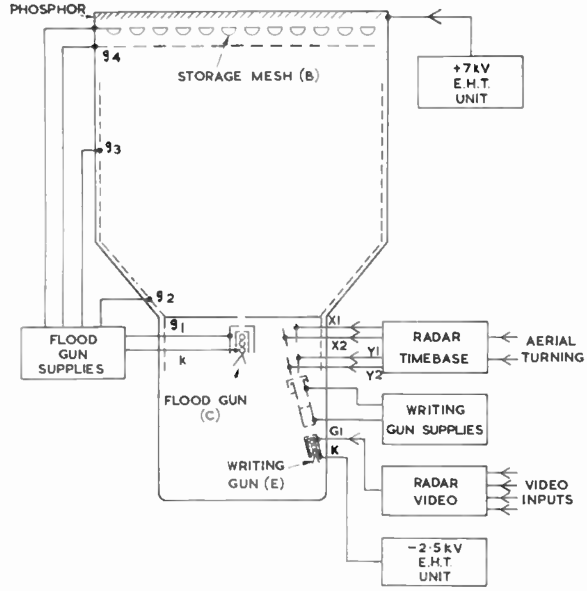


Fig. 4. Basic units of a radar display system using a direct-view storage tube.

deflected, as in a conventional c.r.t. Because of the much higher energy this beam not only reaches the storage mesh, but also causes secondary emission electrons to be released. As the beam moves about the mesh, a positively charged 'trace' is drawn upon it, and flood electrons can penetrate this mesh, reproducing the pattern on the phosphor.

As the potential of any storage element determines the number of viewing beam electrons passing through the holes near the element, a reasonable range of brightness levels can be achieved. The absolute brightness is many hundred times greater than on a conventional c.r.t., and is of the order of 1000 to 2000 foot-lamberts. The stored image can be completely erased by applying a 4-V positive pulse to the backing electrode for about one second. This drives the storage mesh surface positive, by capacitive coupling, and flood electrons can land on the storage surface, and deposit a negative charge until the whole surface is returned to flood gun potential. This completely erases the stored image. When the positive pulse is removed, the storage mesh surface potential drops to -4 V, and cuts off the flood gun electrons. This facility is very useful when altering range or off-centring on a display, as the 'afterglow' is removed, and does not blur the display, as with a conventional fluoride.

By applying the erase pulse for a much shorter period, it is possible to obtain partial erasure. Continuous short erasure pulses enable varying degrees of image persistence to be obtained, control being

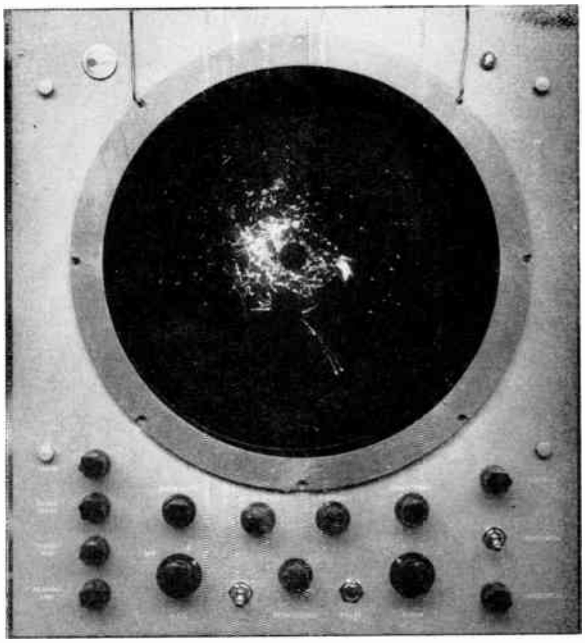


Fig. 5. 'Live' radar on English Electric experimental 11-in. direct view storage tube, under high ambient lighting conditions.

achieved by adjusting the pulse width and/or repetition rate.

The direct-view storage tube is attractive as a radar display partly because it fits into an established system. It needs only an additional e.h.t. unit, and some simple erase circuits. However, so far, the resolution of about 600 lines is only a little over half as good as a conventional display. The storage characteristic is quite different from the afterglow of a fluoride tube. The flash is not present, and the decay characteristic appears more linear. Thus moving aircraft symbols tend to produce rather solid blocks.

Repetitive writing reduces resolution because of charge spreading. For aircraft returns, each echo moves for each successive rotation of the trace, but for permanent echoes or cloud, the addition of successive paints does cause slight spreading. This can be seen on the p.p.i. shown in Fig. 5.

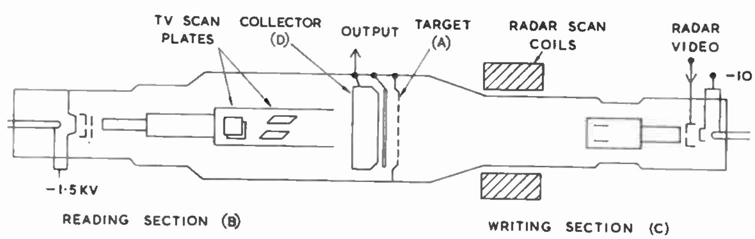
However, for many applications, the direct-view storage tube can provide a useful bright radar display, and future tube improvements may well overcome some of the present limitations.

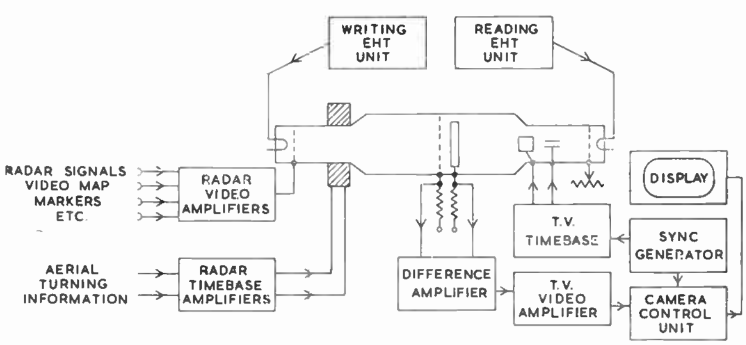
3.2. *Radar/Television Scan Conversion*[2]

The main reason for the low light output from radar displays is the slow basic data rate. One way of increasing the display data rate is by means of scan-conversion, using a bombardment induced conductivity (b.i.c.) tube of the type shown in Fig. 6. This tube is a C.S.F. type TMA 403X.[3] Unlike the direct-view storage tube, it has no visible output, but enables the slowly-rotating radar R, θ time-base to be converted electrically to a high speed X, Y (line/frame) time-base, of a conventional television type. Thus a display as bright as a normal television receiver can be provided.

The target structure A (Fig. 6(a)) consists of a thin sheet of insulator facing the reading gun B. This is normally operated at about earth potential, the writing gun is operated at about -10 kV, and the reading gun at about $-1{\cdot}5$ kV. The radar scanning

(*a*) Scan conversion tube C.S.F. TMA 403X.





(*b*) Radar/television scan conversion system. Block diagram.

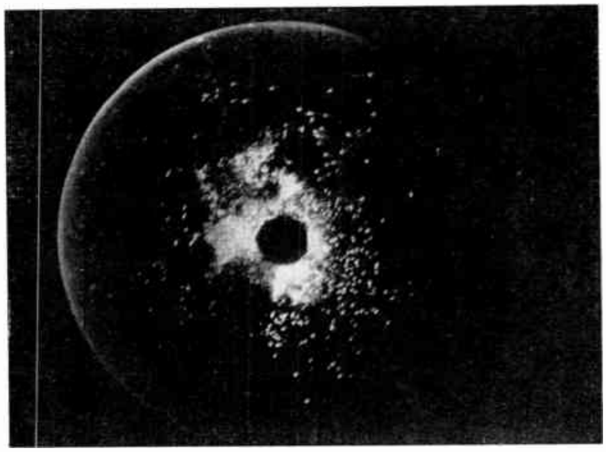
Fig. 6. Scan conversion.

coils control the writing beam by magnetic deflection, and the reading beam 'television' deflection is electrostatic. The collector ring D, from which is taken the output signal, is maintained at a potential of about +50 V. The thin insulator section of the target is the storage layer and is made of a secondary emitting material such as aluminium oxide. The insulating layer also serves as the dielectric of a capacitor formed by the emitting surface of the layer and the metal backing. When the medium velocity reading beam is made to scan the front surface of the target secondary electrons are emitted. The electron velocity is sufficiently high to cause the secondary-emission ratio from this surface to be greater than unity. The target surface begins to charge positively towards the adjacent collector potential and electron emission stabilizes when the potential difference between the collector and surface falls to a few volts.

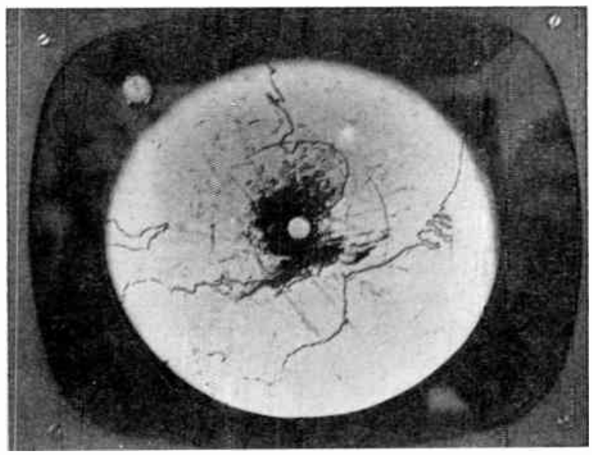
Under these conditions a difference of potential of about 40 V exists between the front and back surface of the target. If the high velocity writing beam is now turned on, it is able to penetrate the target, and induce conductivity through the insulating material. The front surface potential is therefore lowered towards that of the back plate, to a level depending mainly upon the writing beam current. The front surface of the target has traced upon it a relatively negative pattern, corresponding to the written radar signal. Now when the reading beam scans these areas secondary emission will again take place and the emitted electrons will be attracted to the collector ring to provide the output signal current. The output signal is now synchronized with the television scan. By virtue of the secondary emission, the reading beam slowly erases the stored charge image, but, because the charge removed per scan is very small, and because of the large capacitance between the front and back surfaces of the aluminium oxide layer and the high potential difference established between them, a large number of reading scans is required to remove completely the store of image. This gives the afterglow characteristic of this tube. Compared with a conventional radar display, or even a direct-view storage tube, the scan conversion system is relatively costly and bulky. In theory, one scan conversion system can provide signals for a large number of inexpensive television monitors, but in practice, in order to achieve maximum resolution, all the monitors must show the same picture, and this is rarely required. In order to obtain satisfactory resolution, it is necessary to provide a system with about 1000 lines, which is again non-standard and expensive. Furthermore, there is a psychological incompatibility between the rotating radar scan and television line structure.

Perhaps the worst difficulties arise when attempting to integrate a scan-conversion system into a modern air traffic control data handling system. The inflexibility of the system leaves much to be desired.

However, the Federal Aviation Agency in the United States, are providing for scan conversion bright displays for their future air traffic control systems. Scan-converted radar, photographed from television monitors, is shown in Fig. 7.



(*a*) Aircraft trails and ground clutter displayed on television monitor via scan conversion.



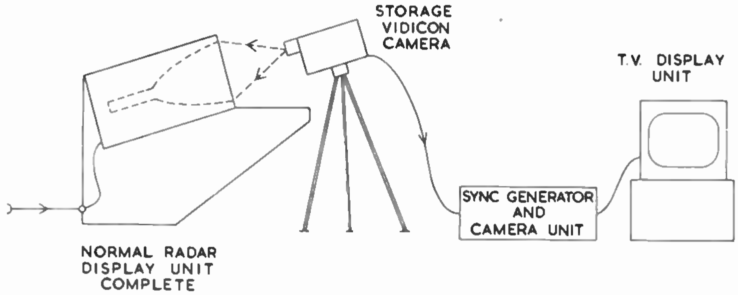
(*b*) A similar picture with inverted television video.

Fig. 7. Scan converted radar.

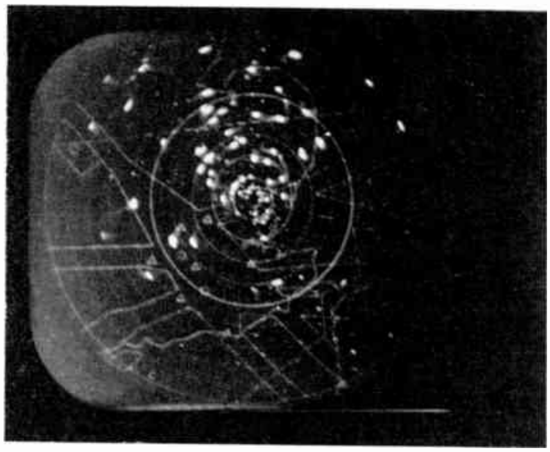
3.3. *Storage Vidicon Conversion*

It is possible to obtain a more flexible scan conversion system than with the b.i.c. tube by using a storage vidicon.

A conventional closed circuit television unit is set up with the camera viewing a normal radar display



(*a*) Radar/television storage vidicon system.



(*b*) 50-cm radar aircraft trails and video map on television monitor, via vidicon scan conversion.

Fig. 8. Storage vidicon conversion.

(Fig. 8). By this means, it is possible to separate physically the radar and television system at the optical transfer point, thereby providing a more flexible arrangement. Unfortunately, for vidicons, long storage appears to be incompatible with good resolution, although several tubes have been obtained with storage times of 3 minutes and resolutions of about 400 lines.

Recent experiments with the Westinghouse 'Permachon' storage vidicon have been very encouraging. Storage times of over an hour are easily obtained, and here the problem is to control storage from, say, 3 minutes to 20 minutes. The most likely method is optically to flash the 'Permachon' target in the television frame blanking period, and experiments are under way to investigate this.

4. Electroluminescence and Cathodelectroluminescence

The phenomenon of electroluminescence occurs when certain phosphors containing suitable activators emit visible light on the application of an external field. If the electroluminescent material is deposited in a thin layer between conducting plates, and one plate is transparent, then a thin plate type of display is obtained.

In order to write discrete and changing information upon the display, the plates must be made up of a multiplicity of parallel conducting wires, perpendicular to each other on opposite sides of the phosphor.

If the number of lines on each side is n then n^2 discrete dots can be displayed. For a resolution of 1000 lines, 2000 connections must be made to the display plate. This is hardly likely to be a practical proposition, even for digital computer controlled systems.

However, recent developments in the United States have included the use of ferro-electric conductors and, more lately, piezo-electric materials.

This has enabled analogue voltages to be applied across the plate, in a manner similar to the use of deflection plates of a c.r.t. However, much work remains to be done to produce a display panel with high speed response time, brightness and resolution, suitable for a bright radar display.

One other interesting phenomena recently reported upon is cathodelectroluminescence.[4] This uses an electroluminescent material inside the face plate of a conventional c.r.t., sandwiched between a transparent conducting surface on the viewing side of the material, and an aluminium backing on the gun side. Each metallic surface is brought out to terminals. It is claimed that the application of a suitable potential across this material controls the afterglow, and that the display is inherently bright.

5. Conclusions

The search for the ideal bright radar display is not an easy one. In the main, the requirement is for a tube similar to conventional fluoride long-persistence display, but several hundred times more bright.

So far, the nearest practical approach to this is the direct-view storage tube. This tube is also capable of development which allows for selective erasure. This facility could be invaluable for computer-controlled systems, as slowly changing data need not be cycled rapidly as is necessary at present to minimize phosphor flicker. The direct view storage tube also allows for negative writing, or 'black-on-white' display. This type of display gives enhanced visibility under conditions of high ambient lighting.

It is obvious that the biggest part of the problem falls on the display tube designer, but unfortunately the commercial market is not a great one and progress is slow. However, there is little doubt that when a bright long-persistence display tube of adequate resolution becomes available at a reasonable price, most ships' officers and airfield controllers will be pleased to have to peer no more into viewing hoods.

6. Acknowledgments

The author wishes to thank the Chief of Research of the Marconi Company for permission to publish this paper. Thanks are due also to colleagues at the Research Laboratories who carried out much of the work described herein.

7. References

1. E. B. Callick and J. C. Fermin, "The principles and applications of storage tubes", *Brit. Commun. Electronics*, **6**, pp. 340–4, May 1959.
2. D. L. Plaistowe, "An experimental scan conversion system for the production of bright radar displays", *Marconi Rev.*, **23**, No. 139, pp. 184–203, 1960.
3. Ch. Dufour, "Les tubes cathodiques a mémoire electrostatique et leurs applications", *L'Onde Électrique*, **36**, No. 355, pp. 801–14, October 1956.
4. *International Electronics*, p. 65, October 1961.

Manuscript first received by the Institution on 28th March 1963 and in final form on 21st June 1963. (Paper No. 868/RNA24.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

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

November 1963

1963 November	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.	1963 November	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	− 130·4	− 131·3	+ 1	16	− 130·2	− 129·8	+ 9
2	− 130·3	− 130·7	+ 2	17	− 130·0	− 130·6	+ 10
3	− 129·9	− 130·3	+ 2	18	− 130·2	− 129·6	+ 9
4	− 129·8	− 130·5	+ 3	19	− 130·5	− 131·6	+ 10
5	− 131·3	− 130·6	+ 3	20	− 131·9	− 130·9	+ 9
6	− 130·8	− 130·0	+ 2	21	− 130·6	− 131·3	+ 11
7	− 130·6	− 130·4	+ 3	22	− 130·2	− 129·6	+ 12
8	− 130·5	− 129·7	+ 4	23	− 131·2	− 130·1	+ 12
9	− 130·7	− 131·6	—	24	− 130·9	− 131·1	+ 12
10	− 131·6	− 131·5	—	25	− 130·9	− 130·3	+ 12
11	− 131·3	− 131·1	+ 6	26	− 130·3	− 130·3	+ 12
12	− 131·2	− 130·1	+ 8	27	− 130·8	− 130·5	+ 13
13	− 130·5	− 129·4	+ 6	28	− 130·6	− 129·8	+ 13
14	− 130·1	− 130·4	+ 8	29	− 130·7	− 130·8	+ 13
15	− 130·6	− 129·7	+ 8	30	− 130·8	− 130·4	+ 13

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 magnetic field.

Note : (1) the phase of the GBR/MSF time signals was retarded by 100 milliseconds at 00 00 UT on 1st November, 1963.
(2) the frequency offset for 1964 will be—150 parts in 10^{10}.

Scientists and Engineers in the O.E.C.D. Area

The Organization for Economic Co-operation and Development has just published a revealing report on the supply of—and the demand for—scientists and engineers in its member countries. The data come from 17 of the 20 O.E.C.D. countries and also from Yugoslavia.. The 300-page report on "Resources of Scientific and Technical Personnel in the O.E.C.D. Area" is the result of a two-year survey, and is virtually the first of its kind.†

It is generally agreed that large-scale increases in the number of scientific and technical personnel are desirable to achieve vast economic growth. Yet little was known up till now on current and future supplies, demands, shortages and deployment patterns of these personnel. The new publication now provides material on which governments and industry may be able to base policies.

A set of definitions for different categories of scientific and technical personnel was first drawn up. These were used as a basis for collecting statistics, thus ensuring a certain degree of international comparability. The Survey data bring out clearly differences in the roles of scientific and technical personnel in the economies of most European countries as compared with American Member countries. On a minor scale, similar differences are found to exist among the European Member countries also. The Survey findings put a number of questions to national authorities concerned with economic, labour market and educational policies:

What are the implications of these differences concerning the functioning of the national economies in question?

Can they be explained by particularities in the structure of national economies, or do they reflect the more or less optimum use of scientific and technical manpower?

Can the present employment of scientific and technical personnel and the expected future development of this employment in the O.E.C.D. Member countries be regarded as adequate in relation to economic growth targets for the present decade?

The Survey data reveal striking differences among individual countries in the composition of the stock of scientific and technical personnel in terms of fields of occupation. Equally significant differences are found in the deployment of the existing stock by sectors of the economy. Here again, the question should be raised whether this wide variety of manpower structures really corresponds to the requirements of the national economies, or whether in some cases existing structures may be an obstacle to economic development.

The stock of scientific and technical personnel in 1959 was 4 to 4·5 million (slightly above 2% of total employment) in the area; 2·6 million were scientists and engineers of university and immediately below university level training. The number of these scientists and engineers showed a notable increase beginning in the 1950's (4% per year) and is expected to continue to increase throughout the 1960's (7% per year in Europe, 6% in North America).

† Publication of the O.E.C.D. Obtainable in Great Britain from H.M. Stationery Office, P.O. Box 569, London, S.E.1.

THE O.E.C.D. COUNTRIES

AUSTRIA	ICELAND*	SWEDEN
BELGIUM	IRELAND	SWITZERLAND
CANADA	ITALY	TURKEY
DENMARK	LUXEMBOURG*	UNITED KINGDOM
FRANCE	NETHERLANDS	UNITED STATES OF AMERICA
FEDERAL REPUBLIC OF GERMANY	NORWAY	
	PORTUGAL*	
GREECE	SPAIN	

* These countries did not take part in the Survey.

In 1959 scientists and engineers represented 1% of total employment in industry, 1·3% in services and 0·1% in agriculture in Europe; while in the United States these proportions were 1·7, 2·6, 0·03% respectively. No marked change is expected in the foreseeable future according to Survey findings.

In 1959, engineers constituted the major part of the total stock of scientists and engineers, 68% for the O.E.C.D. area as a whole (65% in the European and 71% in the American Member countries). The share of natural scientists shows no difference on the two sides of the Atlantic (25%).

In Europe the number of natural scientists is expected to increase in proportion to the total of scientists and engineers during the 1960's. By contrast in the United States the proportion of natural scientists is expected to decline in favour of engineers. The Survey suggests that *shortages* of scientific and technical personnel, though not universal, will continue to be felt throughout the 1960's.

Ample information is provided on the production of scientific and technical personnel from the various national educational systems. If Western Europe's educational efforts, in terms of students and graduates, are behind those of North America, the relevant emphasis within higher education on science and engineering studies is significantly lower in Canada and the United States. During the 1950's the proportion of all students in scientific disciplines to total students in higher education rose from 30% to 34% in European O.E.C.D. countries.

The proportion of first degrees in scientific disciplines to all first degrees in the European Member countries increased from 30% in 1952 to 33% in 1959. Forecasts for some of the European countries indicate 45% for 1970. In the North American Member countries the corresponding proportion increased from 19% in 1952 to 22% in 1959; this percentage is expected to be maintained during the 1960's.

First degrees in scientific studies were obtained by 0·7% of the corresponding age group in 1959 in Europe and by 3·5% in North America. An increase of up to 2% in Europe and 4% in North America is expected by 1970.

While the Survey itself does not contain policy recommendations, the conclusions and analysis will assist the O.E.C.D. countries to judge their own relative position in the context of that of other countries, and to shape their policies accordingly.

U.D.C. 621.396.963: 681.32

The Display of Automatically Processed Radar Information

By

D. R. JARMAN, B.Sc.(Eng.)†

Presented at the Symposium on "Processing and Display of Radar Data" in London on 16th May 1963.

Summary: A system which uses a form of automatic radar data extraction followed by computer processing requires an output device in the form of a display, which accepts digital information and compiles it into a form which can readily be used by a controller.

The display system described in this paper consists of a central equipment and a number of display consoles. The central equipment contains a master timing unit, a buffer store, a symbol generator and a digital/analogue converter, and it generates all the timing waveforms, deflection waveforms and brightening waveforms which are required by the consoles. The consoles themselves are all similar and contain the cathode-ray tube driving circuits together with display selection facilities and a means of injecting information manually into the computer.

The method of operation of the equipment is discussed, and the three basic types of display, and the facilities they offer, are described in detail.

1. Introduction

An essential part of an automatic radar system is the means by which the processed information from a computer can be presented for purposes of monitoring the activity of the system, and to facilitate the exercise of human judgement in control processes. The presentation is usually in the form of a visual c.r.t. display which must provide information separately to a number of individual operators. Thus a complete system will contain a number of display positions.

It is regarded as essential to human appreciation that the display of information from a surveillance radar system should include a p.p.i. type of presentation but, as indicated by Benjamin‡, the picture should not have more information synthetically added to it than the human can absorb without confusion. Hence a display position would contain a p.p.i. picture in which the objects of interest appear as discrete symbol shapes which represent the primary categories of the objects, i.e. passenger-carrying jet aircraft, helicopter, etc.; in addition there would be a separate c.r.t. display on which, in tabular form, the supplementary information available in the system would appear, e.g. speed, height, etc.

Practical considerations set a limit to the degree of automaticity to be provided in such a system, and manual means of entering those data which are not automatically extracted must be provided. It will be convenient to associate this facility of manual injection directly with each display position.

It will also be useful to provide as an alternative the facility to display a raw-radar picture which would enable the system to continue to operate in a limited way in the event of failure of certain parts of the automatic system. An aid under these circumstances is an interlaced controllable marker, by means of which the computer, if it is still functioning, can be kept informed of the most important aspects of the situation.

2. A Complete Display System

In designing a system to meet these requirements, a number of basic facts have to be taken into consideration. There will certainly be a number of controllers using the information provided, and each one of these will need to monitor a different aspect of the situation. One approach to this problem is to tailor each c.r.t. display to a specific operator and provide him with the information he requires. This makes for a very inflexible system, and a very complex equipment for distributing the information to the displays. An alternative approach is to have a standard display unit to which all the information is routed, and to select that which is required at each position. This has great flexibility in that any display unit can be used by any operator, and a central distributing equipment only needs to feed one set of outputs to the displays.

The next question is whether the displays can accept information straight from the computer, or whether some form of buffer storage is required. It is very wasteful of computer time to produce information at a

† Admiralty Surface Weapons Establishment, Portsdown, Hampshire.

‡ R. Benjamin, "Man and machine in the extraction and use of radar information", *The Radio and Electronic Engineer* (*J.Brit. I.R.E.*), 26, p. 309, October 1963.

rate which is suitable for the display system, because, in general, the data rate of the computer is very much greater than this. The repetition frequency of display, too, might be unacceptably low if only a small part of the time of the computer's output circuits were given to read-out to displays. The use of a buffer store is a very desirable feature in that it permits the display to cycle at a conveniently fast rate.

If a system of identical displays is used, then there are two basic approaches to the problem of how to split up the equipment. A system of centralized generation of all waveforms has the advantage of requiring only a small amount of equipment at each display, but has the great disadvantage of the need to feed analogue voltages over long lines. Many of the waveforms required at the displays are timing waveforms and it is perfectly reasonable to transmit these over long lines. However, wherever possible, it is desirable that analogue voltages should be generated individually at each display position.

The injection of information from the display positions to the central equipment must be performed on an individual basis, but even here a common set of lines can be used if a time sharing scheme is adopted. In deciding what principle to use for the injection of information, consideration must be given to the amount of information which has to be passed to the computer, and the number of headings under which it can be grouped. If there are not too many headings, then each specific piece of information can have its own button or switch. However, in many systems, this is not possible, and a more general approach has to be adopted. Under these circumstances, each injection can be set up in the form of a code on a number of standard buttons or switches, which represent, for instance, binary numbers. This is a more versatile system and lends itself to any application, although, in general, more buttons have to be pressed for each injection than in the previous case.

It is useful to be able to point out a target on the p.p.i. display to the computer, in order to ask for more information about this target. A method of doing this is to have a special marker on the display, which can be moved about by the operator by means of a joystick control. Any position indicated by the marker can be transmitted to the computer which can then associate injected information with the marker so specified, or display further information on that marker to the operator.

A simplified block diagram of a system of the type discussed is shown in Fig. 1. All the display consoles in the system are connected in parallel from the outputs of the central equipment.

These principles have been embodied in the experimental equipment to be described.

3. The Central Equipment

The central equipment is shown as part of Fig. 1. The information from the computer is held in the store in digital form and consists of X and Y positional information, heights and categories for the marker display, and symbol information for the tabular display. The category information defines both the category and the marker symbol associated with it. At a time defined by a timing unit (not shown on the diagram), the information for one marker is read out of the store into a temporary store, from which the positional information passes to the digital/analogue converter, the category information passes to the symbol generator and to the category selection at the display, and the height information passes to the

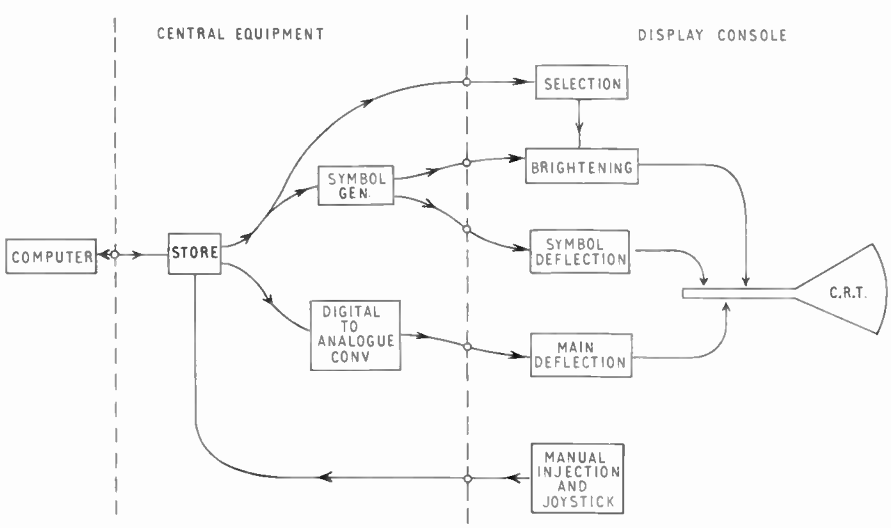


Fig. 1. System block diagram.

height selection at the display. In a similar manner, the tabular information passes to the symbol generator. Another temporary store holds the information from the display units in readiness for insertion into the store. An adder is associated with control of the 'joystick' marker.

The logical operation of the equipment is based on the use of two basic elements. These are the NOR gate and the bistable circuits. NOR gates can be used to perform the basic logic operations of OR, AND and NOT. The bistable circuit can be used as a storage element in a register, a stage of a counter or a delay element.

3.1. *The Master Timing Unit*

The heart of the equipment is a master timing unit which controls the operation of all other units, determining when each other unit is to produce an output, and what type of output it will produce if there is a choice. To avoid confusion, it is not shown on the block diagram. It is basically a sixteen-stage binary counter using the logic units previously mentioned and is controlled by a 500 kc/s master clock pulse derived from the computer; the counting rate is one every 2 µs.

One stage of the counter is illustrated in Fig. 2. By performing logical operations on various combinations of waveforms from the counter, any time interval of 2 µs or multiple of it can be defined. For convenience, the outputs from the counter stages are lettered A to S, where A is the least significant. The two outputs from each stage are labelled output and $\overline{\text{output}}$. (For instance A and $\bar{A}$ in the case of the first stage.)

Certain outputs of the counter are used to produce twelve bits of address for cycling through the store.

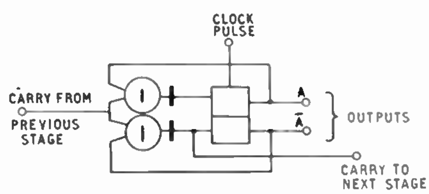


Fig. 2. Typical counter stage.

The store can be controlled by an address produced by the master timing unit or an address provided by the computer. In other words there are two modes of operation, one for extracting information from the store for display, and the other for writing in new information from the computer system. The computer addresses the store sequentially, while the counter addresses the store in such a way that the information required for the various types of display is interlaced in the required sequence. In this way, the most efficient use is made of the time available.

3.2. *The Store*

A ferrite core store is used, with a word length of 24 bits, and a read-write cycle time of 6 µs. The pulses used to initiate the read and write operations are produced by the timing counter.

Three blocks of words are used and each one holds a specific type of information. The first block contains a marker position, category and height for each of the markers required for the plan display. The second block contains the information received from the displays and the third block contains all the necessary information to define the alpha-numeric symbols and their positions in the tabular display.

During operational working of the equipment, the read/write cycle is allowed a total of 16 µs. The period is of this length because the time taken by the symbol writing equipment to produce one marker is of the order of 12 µs, and this has to be produced during one address period. During the time when the computer is feeding information to the display, the read/write cycle time is 6 µs so as to make full use of the speed of the store. This change in timing is achieved in the first four stages (least significant) of the timing counter, which are termed A, B, C and D respectively. Figure 3 illustrates a method of achieving this. When α is a logical 1, the four stages count in the normal way and D has a period of 32 µs, in other words it is counting at the rate of 1 every 16 µs. When α is a logical 0, then the feedback loop from B to A comes into operation and causes one count to be missed. At the same time, the carry from B to C is inhibited and is connected

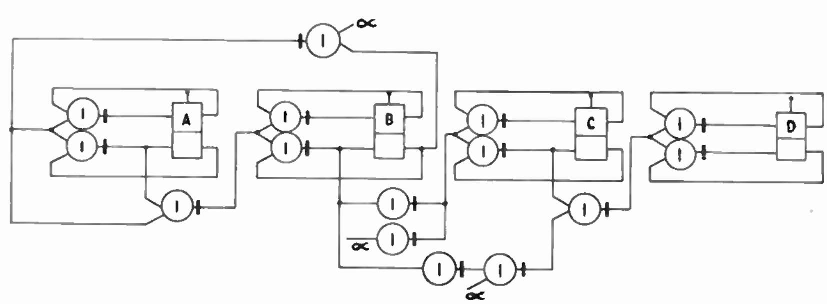


Fig. 3. Four stages of the counter designed to count in increments of 6 µs or 16 µs.

instead directly to D. Thus A and B count up to 3 instead of 4 and D counts at the rate of 1 every 6 μs.

3.2.1. Temporary storage

It is essential for display purposes that the X and Y information for marker positions should be available simultaneously. Since these two are not located in the same store address, at least one of them has to be stored for a short while. In fact, they are both held in temporary stores or registers until they are required by the digital/analogue converters. Information required for symbol generation and selection purposes at the display is also stored in these registers. The registers are composed of bistable circuits which can be set to represent the positional information and the category information.

The joystick marker, which is capable of being moved across the display under the control of a joystick, is made to move by having its positional information 'up-dated' each time it is read out of the store. This 'up-dating' involves adding to or subtracting from either the X or the Y words, or both, a binary 1 during each cycle of the store. It is the joystick setting which determines the co-ordinate and direction of movement.

The accuracy of display is related to the number of bits one is prepared to assign to the X and Y co-ordinates. For instance one might assign 10 bits plus a sign to each co-ordinate, and make the least significant bit represent either $\frac{1}{2}$ mile or less depending on the application. It is unlikely that any military or civil application will require to show more than 32 separate categories and hence 5 bits will be adequate to define category.

3.3. *Digital/Analogue Converter*

In order to make use of the digital positional information held in the store, some form of digital/analogue converter must be employed. To avoid the transmission of analogue voltages, which may be subject to distortion and pick-up, the most suitable form of converter is from digital to time analogue. A time interval can be generated and transmitted with any desired accuracy.

A very useful feature of this type of converter is the ability to change the scale of the display by using a different set of bits with either a lower or a higher significance.

A simplified block diagram of the digital/analogue converter is illustrated in Fig. 4. This shows only one co-ordinate channel for simplicity; the other is similar.

The basic principle of operation is that the 10-bit digital word representing either the X or Y co-ordinate of the marker position is used to set the 10 stages of a counter. The counter is then started and made to count in a reverse direction back to zero.

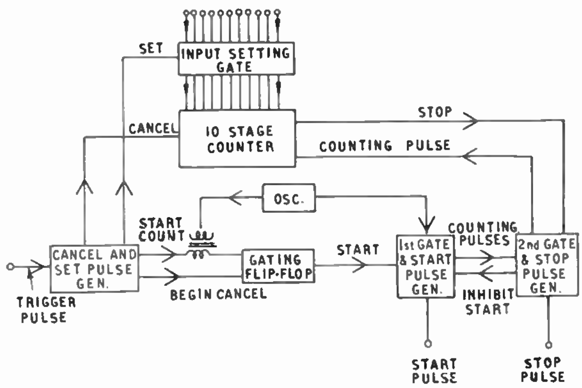


Fig. 4. Block diagram of digital/analogue converter.

When the count reaches zero, the 'all zeros' state is recognized and a 'stop' pulse is produced. The time between the 'start' and the 'stop' pulses is the time analogue of the digital word. The frequency of the oscillator which controls the counter depends on the accuracy required, the maximum range and the time available for a deflection. This last point depends on the response time of the deflection coils at the display. For instance if n mile increments are required to a maximum range of m miles, and the time for deflection is t μs, then a counting rate of $m/n.t$ Mc/s is needed.

The digital word from the register is applied to the input setting gates, and a trigger pulse is applied to the 'cancel' pulse generator. The 'cancel' pulse, lasting a few microseconds, serves to set the counter into the zero state ready to accept the input. At the beginning of the 'cancel', the gating flip-flop is switched over to prevent the counting pulses from reaching the counter. At the end of the 'cancel' pulse, the 'set' pulse is initiated which sets the word on the input lines into the 10-stage counter. The end of the 'set' pulse starts a ringing circuit which has counting pulses superimposed upon it, in such a way that the gating flip-flop is switched off by one of these pulses. This allows the pulses to start operating the counter and to produce a 'start' pulse. If zero is set into the counter, the 'stop' is used to inhibit the start pulse. At the end of the count, a 'stop' from the counter inhibits the counting pulses and produces a 'stop' pulse. The method of synchronizing the 'start' pulse with the counting pulses by means of a ringing circuit is only necessary if a free running oscillator is used. This would not be necessary if a gated oscillator was used which could be arranged to start coherently with the end of the set pulse.

Figure 5 shows a logic circuit designed to produce a square pulse of fixed amplitude, whose duration is controlled by the 'start' and 'stop' pulses. The bistable circuit produces a negative going pulse of the required duration, which is gated through to emitter follower 1

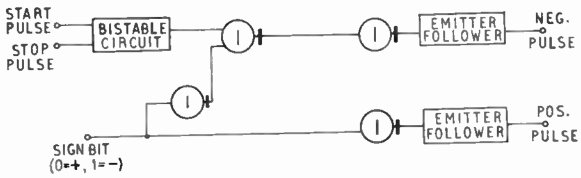


Fig. 5. A logic circuit which produces positive and negative waveforms for deflection.

or emitter follower 2 depending on whether the sign bit is a 0 or a 1. Thus if a positive deflection is called for, e.f.2 produces an output and if a negative deflection is called for e.f.1 produces an output.

3.4. *Symbol Generation*

The symbols which occur on both the marker display and the tabular display are produced by the symbol generator. They are synthesized from a series of 10 straight lines each of which can have a number of fixed lengths and directions and occupies a time of about 0·6 μs, irrespective of amplitude. In fact each one is made up of X and Y components which can have one of four fixed amplitudes.

A simplified block diagram is shown in Fig. 6, and the sequence of operations is as follows.

The trigger pulse switches the Eccles-Jordan circuit and triggers a blocking oscillator (the delay stage), which is included to provide a delay of sufficient length to enable the Eccles-Jordan circuit to reach full amplitude before the gate blocking oscillators start. At the end of the delay period, the 10-gate blocking oscillators trigger in sequence, each one operating from the trailing edge of the previous one. The trailing edge of the last one triggers the Eccles-Jordan circuit back again to its quiescent state. The output of this circuit operates the line pulse generator which feeds a line pulse of approximately 8 μs through the primaries of 28 transformers. There are 32 possible paths through these transformers, each path producing the currents required for a particular symbol. Only one of these lines can be selected at any one time and the selection is achieved by means of a five-bit decoder. These five bits are the five category lines held in the register. Each of the 10 blocking oscillators controls the output from an X transformer and a Y transformer. Blocking oscillators 2 to 9 also control the output from a third transformer which is used to produce the brightening pulse. The reason for stages 1 and 10 not producing brightening pulses is that it is usually most convenient to construct alpha-numeric symbols by starting from a point which is not the centre, and it is therefore made a rule that the first

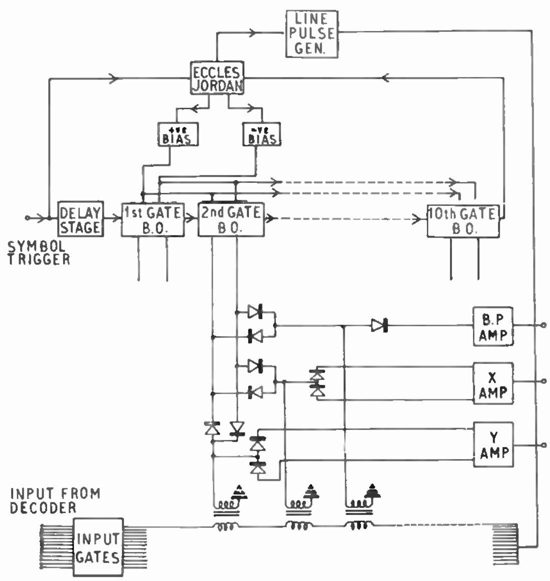


Fig. 6. The symbol generator.

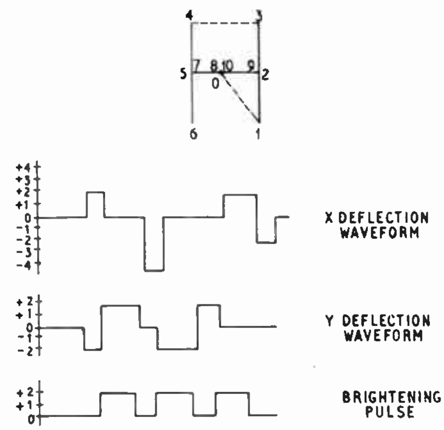


Fig. 7. The symbol 'H' and the waveforms which produce it.

and the last deflections shall not be brightened. The transformers are current transformers all with identical secondary windings, and the four different amplitude levels are obtained by winding 1, 2, 3 or 4 turns on the primary. Under quiescent conditions, the diode clamps on the outputs of the blocking oscillator circuits hold the transformer secondaries at earth potential. When the blocking oscillators operate, the clamp diodes are back biased and the transformer secondaries feed directly to the X, Y and brightening pulse amplifiers. Typical waveforms and the symbol they produce are shown in Fig. 7.

4. The Display Console

A display console consists of three basic units, two of which are identical display units, and the third of which is a panel for the injection of manual information. Each of the display units can be switched to present a marker picture, a tabular display or raw radar. Selection of the tabular display and control of the joystick are effected on the manual injection panel.

4.1. *The Display Unit*

There are four main parts of the display unit, each of which will be discussed separately. These are tube supplies, deflection, brightening and selection. The display is illustrated in block diagram form in Fig. 1.

4.1.1. Tube supplies

The tube used is a 12-in diameter tetrode tube with electro-magnetic deflection. Magnetic deflection has been adopted because, with semiconductor circuits even more than with thermionic valves, it offers a better compromise between brilliance, focus, deflection sensitivity and cost for this type of application. It requires 15 kV on the second anode and 300 V on the first anode. Both of these voltages are produced by a transistor e.h.t. generator which is basically an 8 kc/s oscillator the output of which is stepped up by a transformer to 3750 V peak. This is rectified in a high voltage silicon diode quadrupler circuit and smoothed to produce 15 kV at about 50 μA. The 300-V supply is derived from a suitable winding on the same transformer. The circuit diagram is shown in Fig. 8. The chief advantage of silicon rectifiers over valve rectifiers is that no heater supply is required.

The tube is focused by an electromagnet which requires about 150 mA, stabilized by a transistor circuit against changes in focus coil resistance and voltage.

A note about the type of phosphor used in this tube is probably appropriate at this stage. The choice in this system lies between two phosphors either of which could be used in the displays according to the main requirements. One is a phosphor which produces a bright flash and long afterglow, as conventionally used in raw radar systems, and this is best for use with a marker display where the long afterglow leaves a trail behind a moving marker to provide an indication of movement. The other phosphor considered for use has a relatively flat response with a low level flash, and a much shorter afterglow characteristic. This is most useful for the tabular display where flicker can occur when using the bright flash phosphor. When the tabular information is changed, the old information fades very rapidly and does not become confused with the new information.

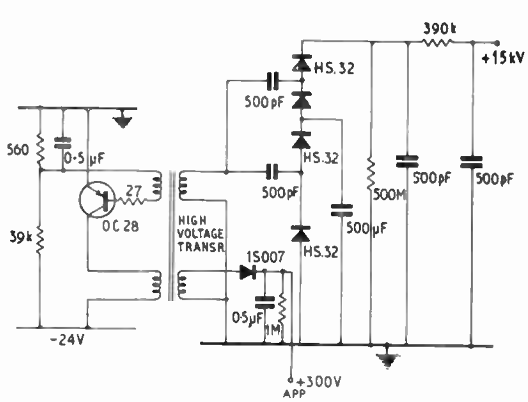


Fig. 8. E.h.t. unit.

4.1.2. Deflection

The deflection system used in the display is electromagnetic, and two separate sets of coils are used, one for deflecting the beam to a given spot anywhere on the tube face, and the other for providing a symbol at that spot.† By this means, it is possible to use a very fast coil for the symbol deflections and so minimize the time required to produce a symbol; interlace problems are also avoided. A further advantage is that the symbols do not increase in size with expansion of the main deflection. The components of the main deflection system are shown in Fig. 9.

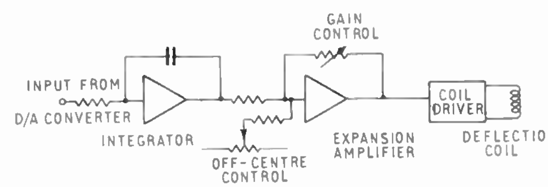


Fig. 9. The main deflection system.

It has already been shown that the digital/analogue converter and associated logic produce either a positive or a negative square pulse which has a fixed amplitude, and a duration which is proportional to the X or the Y deflection. These two waveforms are combined in a buffer circuit which is designed to standardize the positive and negative amplitudes and produces one output for feeding to the integrator. The integrator produces a negative going linearly increasing voltage if the input is a positive square wave, and a positive-going linearly increasing voltage if the input is a negative square wave. At the end of the input square wave, the voltage level reached by the integrator is maintained for a period sufficiently long to enable the current in the main deflection coils to stabilize, and for a symbol to be produced by the symbol deflection coils, before being reset to the zero state. Normal off-centring and expansion facilities are introduced at the amplifier which follows the integrator, and the output of this is fed to the voltage/current converter which drives the coils. The total time allowed for any deflection is 256 μs. Waveform A in Fig. 10 shows a typical input to the integrator, and waveform B shows the reset waveform. Waveform C is the output.

† R. C. Bowes and M. E. Piggott, "A transistor linear time base circuit for a high current electromagnetic deflection", *Proc. Instn Elect. Engrs*, **106**, Part B, Suppl. No. 16, pp. 801–5 and 840, May 1959.

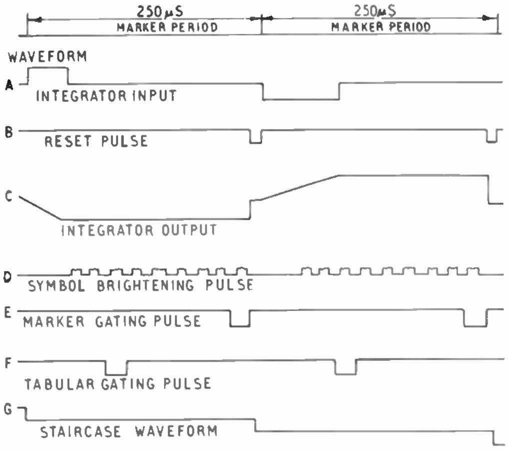


Fig. 10. The timing of the waveforms in the display system.

A similar though less complicated system to that shown in Fig. 9, is used to drive the symbol deflection coils from the input produced by the central equipment.

Waveform G in Fig. 10 shows two steps of a 'staircase' waveform which is used to produce the X-tabular display deflection. The Y deflection is similar in form, though longer in time. Between them, these deflections move the spot into a number of discrete positions, in each of which a symbol can be produced. The integrator is not required in these circumstances, and the staircase waveforms are fed to the X and Y amplifiers. To produce a rotating radar sweep, the inputs to the X and Y integrators from the digital/analogue converters are replaced by square waves of fixed duration, whose amplitudes are proportional to the sine and cosine respectively of the instantaneous angle of rotation of the radar from a fixed reference.

4.1.3. Brightening

The basic principle of this display system is that all the information for one type of display is available at the display units, but only that required to be presented is brightened on a selective basis. Brightening is effected by a combination of two waveforms. One is the marker symbol brightening pulse or raw radar video, either of which is applied to the cathode-ray tube grid, and the other is a suitable gating pulse which is applied to the cathode. The first is positive and the second is negative, and between the two they provide the required beam current for brightening. The brightening pulse or the raw radar video (depending on which display is selected) is available all the time, and selection is achieved by switching the gating pulse. For both the marker display and the tabular display, the brightening pulse from the symbol generator is used. However, the gating pulse in one case is derived from the category and height selection, and in the other case from a selection pulse which decides which of a number of tables is required. The gating pulse for radar is generated to brighten the trace for a given range; at the end of this range, sufficient time is allowed for the spot to return to the centre.

4.1.4. Selection

The selection of markers by height and category is achieved at the display by means of a set of keys. Each marker has associated with it in the store, five bits which represent category, and three bits which represent height.

The incoming category bits are decoded into 32 lines, each of which passes through the contacts of one of the switch positions. When a particular category is selected and a marker appears with this category, a gate is opened and a brightness gating pulse is routed through to the tube cathode via the gating amplifier. Markers are selected by height as those which appear above or below a selected height. In this case, the logic performs the operation of comparing the incoming bits with the height set up on the switch, and deciding whether they are equal and, if not, which one is the greater. Again, the brightness gating pulse is routed through to the cathode if the conditions are satisfied. One of the keys has the ability to override all the selection facilities provided, and to cause all markers on the display to be brightened.

When a tabular display is selected, a different system comes into operation. One of a number of sets of information tables can be selected at the display. These sets of information are written one on top of the other on the face of the tube. In fact, all the symbols required in one particular position are written sequentially on one spot before a further deflection occurs. The selection is achieved by producing a

Fig. 11. Typical marker display.

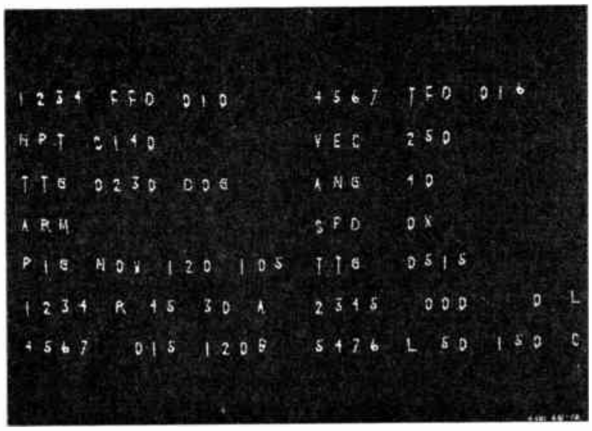


Fig. 12. Typical tabular display.

brightness gating pulse for the required layer. The same pulse selects the same layer from each symbol position and hence the whole table is built up.

4.2. *Injection of Information*

The manual injection and joystick information are both routed to the central equipment over a common set of lines, but a switching arrangement at each display, which is controlled by the master timing unit, ensures that only one display feeds information to the central equipment at any one time. This information is then stored to be read by the computer at a convenient time.

5. The Store Cycle

In order to utilize the time available during a store cycle to its best advantage, and to ensure that each type of display receives information regularly during this cycle, a pair of locations from each of the blocks is addressed once during a basic interval known as a marker period (Fig. 10). The two locations in the tabular display block contain sufficient information to enable eight symbols to be generated, and these are the symbols which appear in one of the defined positions on the tabular display. The two locations in the marker block contain the X and Y information for deflection, and the height and category information for brightening and symbol generation, for one marker on the marker display. The two manual injection locations receive all the information from one display. Thus, while a deflection is occurring on the marker display, eight symbols for the tabular display are being generated, and the manual information is being transferred to the store. When the deflection is completed, a ninth symbol is generated for the marker, and finally, a reset pulse returns the integrator circuit to its quiescent state. The waveforms shown in Fig. 10 illustrate how the interlacing is achieved. This sequence of events is repeated throughout the store cycle.

6. Interlacing Markers with Raw Radar

In the event of a failure of the automatic detection equipment, it is imperative that certain radar echoes should be inserted into the computer manually. This is achieved by interlacing the joystick marker, which normally appears on the marker display, with the raw radar picture. If this can be achieved in the radar dead time (or fly-back time), then no loss of video will result. Unfortunately, however, in this system the radar and the synthetic display are not synchronous, and steps have to be taken to ensure that the joystick marker is available when it is required.

This is achieved in the following manner. When the X and Y deflection information for the marker in question appears at the output of the digital/analogue converter, it is routed via a gate, which is opened at this time, to a second integrator. The second integrator holds its output constant until the next radar reset period when it is added to the radar deflection, which at this time is zero. Having allowed time for the current in the deflection coil to stabilize, a brightening pulse is generated and small sine and cosine deflections are added to the main X and Y deflections to produce a small ring on the display. It will be noted that the symbol generator cannot be used in this case because it is not synchronous with the radar. When the circuits are correctly adjusted, this ring registers with the joystick ring on the marker display. It is quite possible to extend this system to interlace a number of markers with the raw radar picture.

7. Conclusions

A display equipment has been described which acts as the vital link between man and machine in an automatic system. As such, it provides man with all the information he needs in a readily interpreted form.

The standard display console has been designed in such a way that it can be used by an operator or controller in the system, each of whom will need to monitor a different aspect of the situation. Each console contains two display units so that an operator can have both a marker picture and a tabular display. The manual injection panel can be used in many ways, so that the requirements of the system can be met in any possible application.

8. Acknowledgments

This paper is published by permission of the Board of Admiralty. The author wishes to acknowledge the contributions of his colleagues at A.S.W.E. to this work.

Manuscript received by the Institution on 1st April 1963. (Paper No. 869/RNA25.)

U.D.C. 378:621.37/9

Laboratory Work in Advanced Radio and Electronic Education—its Aims and Methods

By

B. F. GRAY, B.Sc.(Eng.) (*Member*)†

A contribution given at a meeting of the Education Group held in London on 23rd May 1962.

Summary: After setting out the aims of laboratory work, the paper discusses the shortcomings of the present methods used in technical colleges. Suggestions are made for types of experiment which call for application of reasoning based on theory rather than merely demonstrate the theory.

This rather awe-inspiring title could be abbreviated to "Laboratory Work—its aims and methods" since there should be no special distinction in the laboratory work in the field of Radio and Electronics. I shall, however, illustrate my remarks largely with examples taken from this field. My remarks will be limited mainly to colleges providing National Certificate courses, although many points will apply to any institution. For this reason the project type of experiment is outside the scope of my paper.

Laboratory work—what are its aims? Are these aims worthwhile? Is laboratory work in fact absolutely necessary? Different courses and different institutions place differing emphasis on it. It is necessary in the case of a National Certificate student, for example, for at least one-third of his total time to be devoted to laboratory work. This proportion of the time is probably larger than that demanded by some universities.

Is then this expenditure of time and money on the practical aspect a good thing? Should certain universities devote more of their time to practical work at undergraduate level? Do the technical colleges spend too much time in laboratory work? Is the form taken by the laboratory work suitable?

Before attempting to answer these questions let us start by stating the aims of laboratory work which come to mind.

1. It is a practical illustration of some theoretical deduction which may or may not be fully understood. That is, it may be used to verify a well-established theory or to verify some new theory. This is an important difference. The former lies more in the undergraduate field, the latter more in the research and postgraduate field.
2. It is a means of finding out something not already known. Again, this could be at any level. A very simple illustration would be to find the value of an unknown capacitor.
3. It is a means of finding out the practical limitations of some theoretical deduction.
4. It is a method of training in techniques of measurement.
5. It enables students to become familiar with apparatus and equipment.
6. It can be a teaching aid.
7. It can become a method of training in logical deduction.
8. Finally, in the written report, which is a common adjunct of the majority of laboratory work, it calls upon students to present results and records and to make observations and conclusions which must be understood by others. It is thus an exercise in communication.

Now, in general, I believe that laboratory work is often very badly carried out. Many students regard the practical period as merely a prelude to a written report and if given the opportunity would be only too glad to give up laboratory work completely. In many technical colleges students are told that they must perform ten experiments per subject and this exercise is carried out once every three weeks. Let us imagine what goes on.

The students have been divided into groups of three or more and are spending the first half-an-hour or so of the laboratory period copying out word for word the instruction sheet for the experiment placed in front of them. The well-organized group we are watching has elected a chief scribe who produces carbon copies for his colleagues. The chief scribe is quite content to act later as chief recorder of results (again with carbon copies) and it is doubtful whether he has bothered to examine the equipment in front of him. The other members of the group appear quite keen to carry out the experiment (detailed instructions being provided) and after the usual delay in which

† Head of Department of Electrical Engineering, Hatfield College of Technology, Hertfordshire.

equipment is connected together, circuits are checked by a member of the teaching staff and a series of values of current and frequency, voltage, etc. are recorded. It is often difficult due to inevitable delays to complete the experiment within the allotted time and, consequently, this group is very pleased to find that it has finished before the end of the period. The disappointment of course comes later when they find that a most important parameter which should have been maintained constant was allowed to drift and, in consequence, the results are not conforming to theory. Fortunately, Joe Bloggs carried out the experiment last week and his results followed precisely the theoretical pattern so that the results may be borrowed and all is not lost. Three good but rather similar reports ultimately are submitted to the lecturer (Joe Bloggs got 8 out of 10) and three more students have qualified in coursework.

It is worthwhile considering how much of value was obtained from this exercise. Certainly the importance of a particular parameter was brought home but this was achieved after the experiment was a failure although in each report the experiment was of course—to quote a phrase—'considered a success!' The scribe in this group incidentally may not even have appreciated this point.

Now it is so easy to improve on this technique. One can stop the waste of time at the start of a lecture period by issuing instruction leaflets—preferably before the laboratory period so that the keener type of student may have thought about the experiment *before* carrying it out. But one must get at the deeper implications of this fairly typical story.

Students were more prepared to accept the theoretical deduction than their observed results. They questioned their results and not the theoretical work. In that case would it not be far better to use the experiment as a practical demonstration of the theory either during the lecture or immediately afterwards. There is nothing to prevent individuals if they wish taking their own results and plotting their own graphs, but it would seem largely pointless to allow each student to set up his apparatus some time later in the session merely to verify some theoretical work already accepted. The only case for such an action would lie in the revision or re-emphasis of some aspect.

Thus, under heading (1) the student may derive more from good demonstrations than from actual experiments.

There seems to be a much stronger case for making item (2) the main purpose of laboratory work. Too often one hears the plea—we cannot do this experiment, we haven't covered the theory. In many cases—why cannot the experiment be performed? A rather different approach may be required but at least students are using practical work in a sensible manner and often respond very favourably to this more adult approach. Certainly no one can be quite certain how the various parameters will vary and there is far more reliance placed on one's own results. The experimental work has more meaning, more importance, and there is incentive to use the library services to find out something about the experiment. Let us take a simple illustration of this.

A well-known experiment involves the measurement of the potential drops across two components—a resistor and capacitor in series—as the frequency of a constant voltage applied across them is varied. Students are often told that the vector sum of these voltages gives the supply voltage and they are shown how to obtain the locus of the current.

When this experiment is conducted by the student, they will make quite certain that their results conform to theory and a perfect semi-circle results. I wonder whether they observed that initially at low frequencies the p.d. across the resistor increased linearly with frequency and at high frequency there was little variation in this voltage—no, of course not, they were so absorbed in obtaining a semi-circle that the frequency response of the circuit was missed.

It is easy to rearrange this experiment. Assuming that the basic elementary a.c. circuit theory has been covered (and here is a case where some theory must precede the practical), give the students a black box with two terminals and ask them to identify the components in the box using a multi-range meter and a wide-range oscillator. Only those who have seen students carrying out this type of experiment will appreciate the increased interest shown.

Another example where the experiment could with great benefit precede the theory is the maximum power transfer theorem. Give the students a black box containing a battery in series with, say, a 100 Ω resistor. Ask the students (i) to find the internal resistance using merely a decade resistance box and a multi-range meter, and (ii) to find the conditions for maximum output power. Again, there is considerable enthusiasm in carrying out this experiment and later the theoretical deduction of the maximum power transfer in the lecture period is quite superfluous. Moreover, the keener students will even take into account the impedance of the meter which is not completely negligible (deliberately so). The spectrum of ability of students carrying out this type of experimental work is established very early in the session. It will be necessary of course to give the students ample notice of such experiments and it is desirable to provide a tutorial class before the laboratory period.

And now a word regarding the written report. I find no purpose whatsoever in asking students to produce a formal report of each experiment carried out. Wherever possible, I ask students to maintain a laboratory logbook which is not marked for neatness but merely on results obtained, observations made and conclusions reached. The students are then asked to produce roughly one-third of the experimental work in a formal manner. The formal report is marked of course with an eye on the grammatical presentation. I feel this approach is a desirable compromise.

Some work should be produced formally to instill in the students a certain pride of presentation in their work and also to bring home to them the fact that the results of their experimental work may have to be read and understood by others not necessarily engaged in the field of electrical engineering. An even better exercise at this stage is to ask students to comment on their experimental results before a lay audience and answer questions. This can normally be achieved only with full-time students due to the lack of time available with part-time courses.

It may be a little surprising to teaching staff using this approach for the first time to find that not all students are in favour of this attitude to the formal report. Some would prefer to produce all their work formally and dispense with the logbook. Such students declare that the logbooks make work—they have to make a fair copy in their logbooks and then in some instances produce a formal report in addition. These students are nearly always the type who are good draughtsmen, are capable of producing a remarkably neat report but incapable of coming to a reasonable conclusion. One can only hope that these people will understand some of the implications of laboratory work and not regard it as a means of demonstrating their draughtsmanship. Only too frequently do these students gain high coursework marks but fail to pass the examination.

The other points mentioned at the beginning are largely self-explanatory—of course, laboratory work is a method of training in measurement techniques and enables students to become familiar with equipment and this does imply that the laboratory equipment should not be too standardized and the method of measurement should not be too stereotyped.

Item (6) on the list is a point which is seldom made: practical work may be used as a teaching aid. This is intended to have a rather subtler meaning than the use of demonstration apparatus. Demonstrations are excellent but suffer from the big disadvantage that they are carried out by the demonstrator and not by

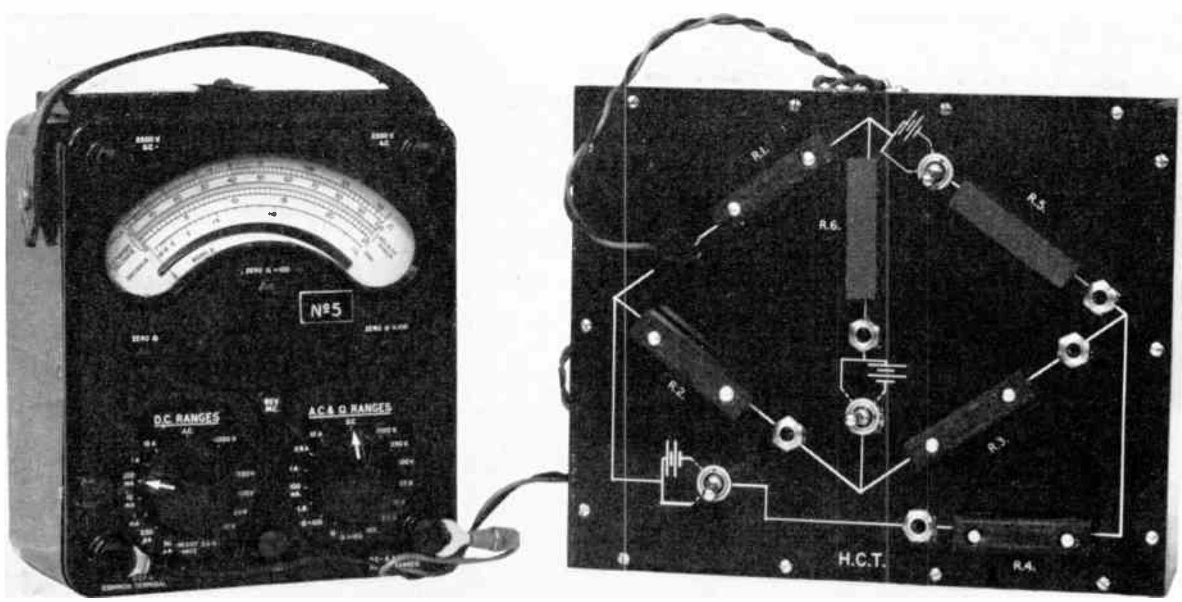


Fig. 1. Variable-configuration circuit box for experiments in the laboratory. Three sample plug-in components are shown on the right.

the student. They can replace poorly organized practical work but not work that brings out the majority of other points I have raised. Laboratory work which is carefully planned with the correct aims always in mind can never be replaced by demonstration. But the use of certain equipment at a critical point in the development of some aspect of the theoretical work can do a great deal of good. For example, let us take a lecture on network theory. The discussion may have centred around certain theorems —Superposition and Thévenin, which are difficult to understand. The brighter students may be fairly happy with this topic but the duller ones are still having difficulty especially with the Thévenin idea of "the resistance looking back in the circuit with all e.m.f's replaced by short circuits". If a black box is produced in the classroom at this stage which any student can play with, how much easier it is to grasp this idea. The box shown in Fig. 1 may be used for demonstrating a variety of experiments, and verifying a number of circuit theorems.

Conclusions

Laboratory work as carried out in many technical colleges often lacks purpose and is consequently not regarded by the students as a necessary adjunct to the theoretical work. It requires only a little thought to change much of the existing laboratory work into something which is a vital part of a course of study.

It is not essential in many cases to cover the theory before the practical work and in some instances the laboratory work is enhanced by dealing with the theoretical work after the experiment. The timing of the laboratory work is important. Good demonstrations may achieve more than some experiments at undergraduate level but cannot replace the carefully thought-out experiment. Students respond much better to the 'find-out' or problem type experiment than to the verification kind of practical work and a much better indication of the progress of an individual may be found from this approach.

As a teaching aid comparatively little use is made of the practical side at present but much useful development could be carried out in this direction. There should be a great deal more research into the whole matter of practical work and we should certainly not be slaves to convention.

The time spent in laboratory work can be used much more efficiently than at present and the students can derive a great deal of educational benefit from it, although the way laboratory experiments are presented to students may have to be varied for different calibres of students.

Lastly, the written formal report serves a useful but limited function, and students should be encouraged to adopt the logbook method of recording results. The value of a formal report is enhanced if it is presented to a lay audience who can ask questions about the content of the report.

Bibliography

E. M. Williams and F. J. Young, "Electrical Engineering Problems", (McGraw-Hill, New York, 1960).

B. F. Gray, "Electrical laboratory work—one-third of the time", *Technical Education*, 3, p. 30, January 1961.

P. F. Soper, "The philosophy of laboratory work", *J. Instn Elect. Engrs*, 8, p. 251, May 1962.

Manuscript received by the Institution on 30th July 1962. (Contribution No. 77/Ed.12.)

U.D.C. 621.396.677

Non-collinear and Cylindrical Multiplicative Arrays

By

E. D. R. SHEARMAN,

B.Sc.(Eng.)†

Summary: It is shown that an array, in which the element outputs are multiplied together in pairs and the smoothed products are added, has a directional pattern independent of whether or not the various pairs are collinear. This property may be useful in sonar and radar systems in easing the physical requirements of an array. In particular, two circular scanning arrays are proposed, and their directional performance is compared with that of the corresponding additive array.

1. Introduction

There has been a great deal of discussion in recent literature[1-5] of the performance of various array arrangements in which signals received on two or more portions of the array are multiplied together instead of being just added. Generally it has been found that the multiplicative arrangements give narrower directional patterns (defined as the variation of output voltage of the signal received from a single small source in the far-field as the array is rotated about its centre) than the same array used additively, together with some other advantages and some disadvantages. However, it appears that what may be a rather important advantage of a certain type of multiplicative array has not been generally recognized, namely, that the array need not be collinear or uniform to give the same directional pattern as a collinear and uniform array. It must be admitted that this is implicit in the particular aperture-synthesis technique used by Ryle[6] in radio-astronomy, where it permits a reduction in the number of positions which the movable pair of elements has to occupy; but it is felt that an explicit statement is desirable, since there are some more obvious applications.

2. Collinear and Non-collinear Arrays

Consider a collinear array containing an even number of elements, as shown in Fig. 1, in which elements 1 and 1′ are spaced by a distance d, elements 2 and 2′ by $3d$, elements 3 and 3′ by $5d$, etc. The output signals from 1 and 1′ are multiplied together and smoothed; similarly for 2 and 2′ and for 3 and 3′, and so on. All the smoothed outputs are added together. For a signal of the form $\cos(\omega t+\alpha)$ received at the centre point from a distant source lying in a plane containing the line of the array, the signal outputs of each pair of elements spaced a distance rd apart are of the form $\cos(\omega t+\alpha-r\phi)$ and $\cos(\omega t+\alpha+r\phi)$, where ϕ is determined by the direction (θ) of the received signal, i.e. $\phi = (\pi d/\lambda)\sin\theta$, where λ is the wavelength. The sum of the smoothed outputs is therefore of the form

$$\sum_{r_{\text{odd}}} E_r^2 \cos 2r\phi \qquad \ldots\ldots(1)$$

where E_r is the output voltage from each of the two elements spaced by rd. If all E_r are equal, this becomes proportional to

$$E^2 \frac{\sin 2(m+1)\phi}{\sin 2\phi}$$

where m is the highest value of r. This is recognizable as being a directional pattern of the same shape as obtained with an ordinary collinear additive array of the same geometry, but is of half the beamwidth.

Now it should be observed that the phase angle α, which represents the phase of the signal as received at the centre-point of the line joining each pair of elements, plays no part in the process, which is concerned only with the phase relations within each pair of signal outputs. Thus, provided the signal is c.w., or of adequate duration, it is not necessary for the pairs of elements to lie on the same line, provided the axis of each pair is parallel to the others. The arrangements of Fig. 2, (a), (b) and (c), therefore all give exactly the same directional response. Moreover, there is no need for the centre-points of the lines

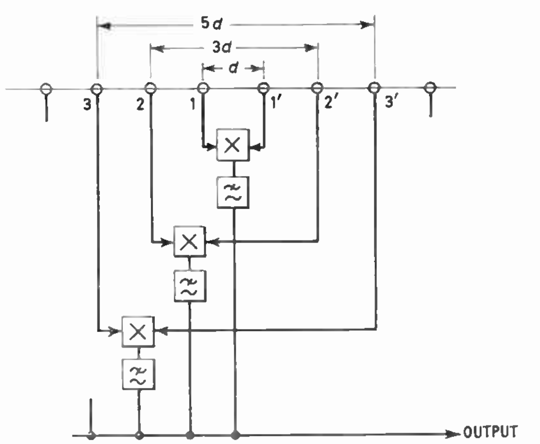


Fig. 1. Multiplicative array.

† Electrical Engineering Department, University of Birmingham.

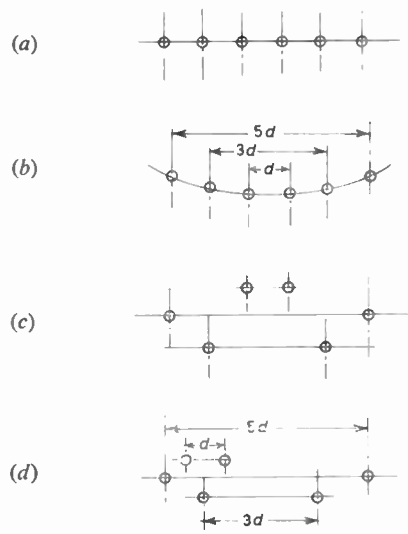


Fig. 2. Arrangements of six elements all giving the same directional pattern when used multiplicatively according to Fig. 1.

joining the pairs to be coincident, so that the arrangement of Fig. 2 (*d*) also gives the same directional response. This rather surprising result is not obtainable with ordinary additive arrays, and may well be of some benefit in practice. For example, a large sonar array might be made of transducers mounted at intervals on the hull of the ship without the shape of the hull having to be taken into account.

3. Steerable Cylindrical Arrays

The property discussed above may also be valuable in the design of a circular or cylindrical array to fulfil the function of the p.p.i. sonar system[7] or the 'Wullenweber' high-frequency radio direction-finder.[8] In these systems a receiving system has to be continually swung through all bearings. A cross-section of such an array is shown in Fig. 3, where 16 elements are used. For multiplicative operation, a rotary contact-type or capacitance-type of mechanical switch, or an electronic switch of the same type, is used to connect the elements in multiplicative pairs, thus:

First position: 1–1′, 2–2′, 3–3′, 4–4′, 5–5′, 6–6′, 7–7′, 8–8′.

Second position: 1′–2′, 1–3′, 2–4′, 3–5′, 4–6′, 5–7′, 6–8′, 7–8.

Third position: 2′–3′, 1′–4′, 1–5′, 2–6′, 3–7′, 4–8′, 5–8, 6–7 and so on.

The dimensions of, and number of elements in, the array would need to be arranged so that each position of the switch moved the beam no more than one beamwidth. For rapid scanning one complete rotation should be achieved within a time equal to the duration of the transmitted pulse (assuming an active echo-ranging system) or within a time equal to the reciprocal of the signal bandwidth at the array (assuming a passive 'listening' system).

The directional pattern of the multiplicative cylindrical array in a plane containing the cross-section shown in Fig. 3, and assuming equal sensitivity on all elements, is easily shown to be

$$\sum_{m=1}^{m=n} \cos 2\phi_m \qquad \text{......(2)}$$

where the array comprises $4n$ equally-spaced elements on a radius R, and

$$\phi_m = \frac{2\pi R}{\lambda} \sin\left((m-\tfrac{1}{2})\pi/2n \right).\sin\theta \qquad \text{......(3)}$$

Here θ is, of course, the direction of arrival of a wave relative to an axis normal to the lines of the multiplicative pairs. For the 16 element array shown in Fig. 3, this directional pattern normalized to a peak height of unity is shown in Fig. 4. A diameter/wavelength ratio of 0·9 has been chosen so that the pattern advances one half-power beamwidth every time the rotary switch is advanced one position (360°/16 = 22·5°).

An ordinary cylindrical array[7] (i.e. with all elements connected additively through phase-compensating circuits) has a somewhat similar, but wider, main beam and first side-lobe, as shown by the dashed-line curve in Fig. 4. The derivation of the directivity expression for the phased additive array is given in the Appendix.

The multiplicative array has, of course, a d.c. output, so that the negative polarity of the main secondary responses means they either do not cause an indication on the display, or can be eliminated (for single targets) by the use of a rectifier. Apart from this and the narrower beamwidth, the big advantage of the multiplicative arrangement would clearly be the elimination of phase-compensating circuits; this represents in many applications (as in the p.p.i. sonar) a very large saving of weight and cost.

A disadvantage, evident from Fig. 4, is that the pattern has a unity front/back ratio. It is possible

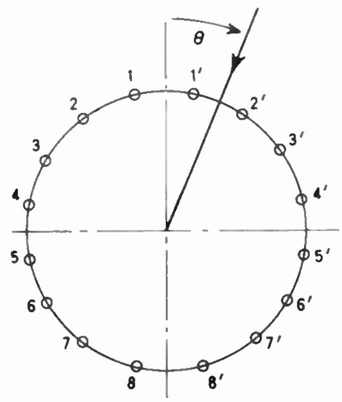


Fig. 3. Sectional diagram of cylindrical array.

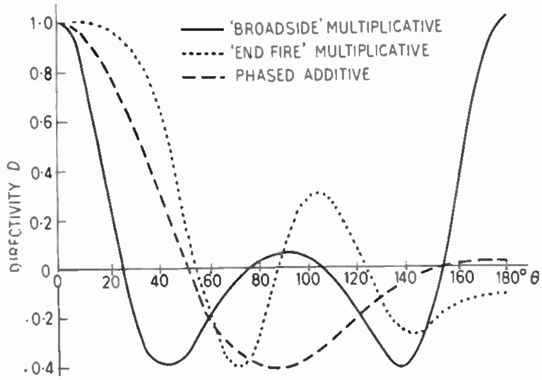


Fig. 4. Directional patterns of circular arrays.

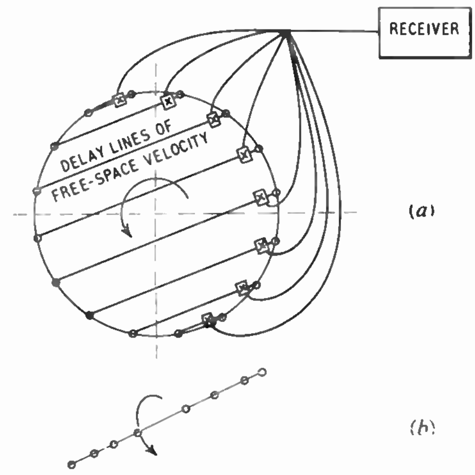


Fig. 5. End-fire multiplicative array.
(*a*) Schematic of array.
(d) Equivalent linear rotating array.

that this might not be important in some applications where the advantages might outweigh it. Moreover, if search over 180 deg is really all that is required (as may be the case in many practical applications) then the difficulty may be mitigated if each element of the array is made to have an individual directional pattern covering approximately 180 deg. This is not difficult with electro-acoustic transducers and some microwave radio aerials, although shadowing of one element by another may occur unless there is some staggering of the elements in the axial direction; an alternative is to use end-fire elements.

The freedom from backward radiation in the phased array is due to the end-fire action of the pairs of elements (see Appendix). This end-fire property is necessarily lost in the process of multiplication. End-fire characteristics can be obtained, however, if a delay line with delay corresponding to direct wave travel from elements 1 to 1′ is inserted in the output of 1 before multiplication with the undelayed output of 1′ (with corresponding arrangements for the other pairs). The physical arrangement (shown in Fig. 5 (*a*)) is equivalent to the rotating end-fire linear array shown in Fig. 5 (*b*). The directional pattern of this arrangement is easily shown to be

$$\sum_{m=1}^{n} \cos 2\psi_m \qquad \ldots\ldots(4)$$

where, again, the number of elements is $4n$, equally spaced on a radius R, and

$$\psi_m = \frac{2\pi R}{\lambda} \cos\left[(m-\tfrac{1}{2})\pi/2n\right].(1-\cos\theta) \qquad \ldots\ldots(5)$$

For the same diameter/wavelength ratio as chosen above, the directional pattern of a 16-element array of this type is shown by the dotted-line curve in Fig. 4. The width of the main beam is greater, and a larger diameter/wavelength ratio is clearly needed, but the front/back ratio has been improved to 9 : 1.

For radio aerial systems, unlike sonar, the introduction of the delay lines is straightforward and this method offers a possible technique for obtaining back/front discrimination.

4. Conclusions

A study of the directivity expression of a collinear array in which element outputs are multiplied in pairs and the smoothed products are added, shows that the mean r.f. phase of each pair has no effect on the output. This suggests that the pairs may be non-collinear without affecting the pattern, a property which has already been used in aperture synthesis in radio astronomy.

It has been shown that this principle could be usefully applied to sonar and radar systems to obtain a more physically convenient array arrangement. In particular two circular scanning arrays are proposed.

5. Acknowledgment

The author wishes to thank Professor D. G. Tucker who suggested this investigation.

6. References

1. A. Berman and C. S. Clay, "Theory of time-averaged-product arrays", *J. Acoust. Soc. Amer.*, **29**, p. 805, 1957.
2. V. G. Welsby and D. G. Tucker, "Multiplicative receiving arrays", *J.Brit.I.R.E.*, **19**, p. 369, 1959.
3. V. G. Welsby, "Two-element aerial array", *Electronic Technology*, **38**, p. 160, 1961.
4. V. G. Welsby, "Multiplicative receiving arrays: the angular resolution of targets in a sonar system with electronic scanning", *J.Brit.I.R.E.*, **22**, p. 5, 1961.
5. D. G. Tucker, "Multiplicative arrays in radio-astronomy and sonar systems", *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*), **25**, p. 113, 1963.
6. M. Ryle, "The Mullard radio astronomy observatory", *J. Instn Elect. Engrs*, **6**, p. 14, 1960.
7. J. W. Horton, "Fundamentals of Sonar" (U.S. Naval Institute, Annapolis, 1957.)
8. C. Crampton, "Naval radio direction-finding", *J. Instn Elect. Engrs*, 94, Part IIIA, No. 11, p. 132, 1947.

7. Appendix

The Directional Pattern of a Phased Additive Circular Array

Consider a circular array, such as that in Fig. 6, which is to be arranged so that the outputs from all elements add in phase at the receiver input when a wave is incident at $\theta = 0$. If cable of free-space velocity is available, the phase compensation arrangement shown in the figure is suitable.

For the element m in Fig. 6, the phase retardation in the cable is

$$\frac{2\pi}{\lambda}(l_1 + l_2 + R\cos\alpha_m)$$

For a wave at $\theta = 0$ giving a field A cos pt at 0, the field at m will be

$$A\cos\left(pt + \frac{2\pi}{\lambda}R\cos\alpha_m\right)$$

The signal at the receiver will thus be proportional to,

$$A\cos\left[pt - \frac{2\pi}{\lambda}(l_1 + l_2 + R\cos\alpha_m) + \frac{2\pi}{\lambda}R\cos\alpha_m\right]$$

$$= A\cos\left[pt - \frac{2\pi}{\lambda}(l_1 + l_2)\right]$$

This is clearly independent of α_m, so that the contribution from all the elements at the receiver input will add in phase as required.

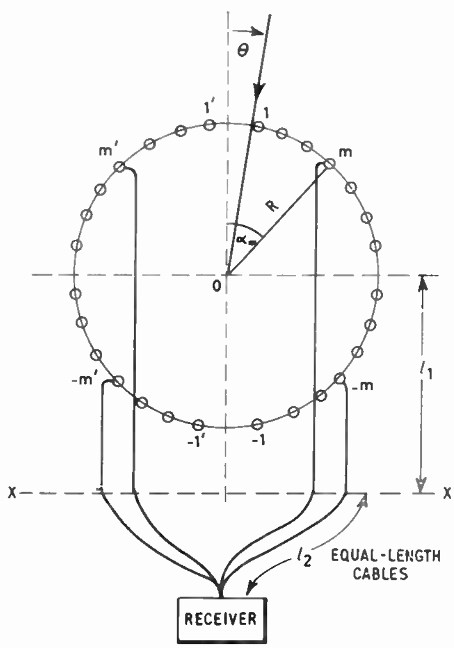


Fig. 6. Cable arrangement to obtain phase-addition from a circular array of elements from a wave incident at $\theta = 0$ deg. (Free-space velocity for the cable is assumed.)

For a wave incident at $\theta > 0$, the phase retardation in the cable will be unaltered, but the field at the element m will now be

$$A\cos\left\{pt + \frac{2\pi}{\lambda}[R\cos\alpha_m\cos\theta + R\sin\alpha_m\sin\theta]\right\}$$

The signal at the receiver input is therefore proportional to

$$A\cos\left\{pt - \frac{2\pi}{\lambda}[l_1 + l_2 + R\cos\alpha_m] + \right.$$

$$\left. + \frac{2\pi}{\lambda}[R\cos\alpha_m\cos\theta + R\sin\alpha_m\sin\theta]\right\}$$

Clearly the phase compensation is now incorrect, and there will be phase differences between the contributions from the various elements.

It is convenient to consider the vector summation of the contributions in pairs, taking m, m' together and $-m$, $-m'$ together. Both pairs give resultants having the same 'broadside' directional pattern. However, when the resultants of the two pairs are added, the phase difference between them results in an end-fire directional pattern which is multiplied by the broadside pattern.

This process leads to a directional pattern for the complete array of $4n$ elements of the form

$$D = \tfrac{1}{4}\sum_{m=1}^{n} D_1 . D_2$$

where $$D_1 = \cos\left[\frac{2\pi}{\lambda}R\sin\alpha_m . \sin\theta\right]$$

and $$D_2 = \cos\left[\frac{2\pi}{\lambda}R\cos\alpha_m . (1 - \cos\theta)\right]$$

in which $\alpha_m = (m - \tfrac{1}{2})\dfrac{\pi}{2n}$

The first term in the product represents the broadside directivity of an element pair such as m, m' and is symmetrical about $\theta = \pi/2$. The second term represents the end-fire directivity of the two pairs of elements such as m, m' and $-m$, $-m'$ and is not symmetrical about $\theta = \pi/2$.

It is interesting to notice that the first term is identical (apart from a factor of two due to additive instead of multiplicative processing[2]) to eqn. (2) obtained for multiplying in broadside pairs. Similarly the second term is identical (again apart from the factor of two) to the eqn. (4) obtained for the end-fire multiplicative system of Fig. 5.

Manuscript received by the Institution on 13th February 1963 (Contribution No. 73/RNA26.)

U.D.C. 621.375.018.756: 681.3

High Speed Logic Circuits using a Tunnel Diode Transistor Feedback Amplifier

By

G. G. SCARROTT†

AND

R. W. MITCHELL, M.A.†

Presented at a Computer Group Symposium on "Tunnel Diodes as Switching and Storage Elements" in London on 23rd January, 1963.

Summary: A fresh approach is made to computer circuitry in which due prominence is given to the electrical characteristics of the logical interconnection network. The necessity for matching active logic elements to a screened transmission system rules against the use of diode-gate transistor amplifier circuits of the type which has become familiar.

An amplifier circuit is described which has a virtual-earth input combined with a non-linear response suitable for threshold logic. This circuit exploits the high-speed amplifying and standardizing properties of a tunnel diode in conjunction with the ability of a transistor to direct and buffer information flow. Unlike most hybrid circuits, the transistor is allowed to give current gain without loss of speed. The action of the amplifier is inverting and latching, i.e. it is bi-stable. The degree of latching may be controlled by the use of delayed negative feedback, and other useful variations are possible.

Preliminary experiments have been carried out with slow tunnel diodes, the circuits being suitable for 30 Mc/s operation if used in a synchronous mode.

1. Introduction

The information rate of present-day logic networks is limited as much by transmission time between active elements as by delays in the elements themselves. Reduction in physical size in proportion to signal 'wavelength', as may be realized with integrated circuit techniques, is not a complete answer to this problem, particularly since the fastest active circuits may not prove integratable. By the use of screened logical wiring with matched terminations it should be possible to reduce and control transmission delays, and to improve overall noise rejection. The design of new computer hardware may include planned departures from this principle, but adherence should be the rule.

Faster active circuitry must retain the essential attributes of power gain and non-linear response. Low gain restricts speed in the divergent parts of the network, and makes the system sensitive to losses. Non-linearities should be as sharp as possible in order to reduce uncertainties in signal amplitude, an ideal characteristic for a neutral logic amplifier being shown in Fig. 1.

The introduction of comprehensive impedance matching immediately gives a scale to the quantities involved, since transmission lines of practicable impedance must be included in the system. Hence, the signal limits defined by the break-points of the non-linearities to be employed will, in turn, define the power level at which the amplifiers must operate. Printed strip lines on circuit boards can cover roughly a decade of impedance values between 30 and 300 ohms, while commonly available coaxial cables lie in the narrower range of 50 to 125 ohms. Both types of line can be designed to provide adequate screening.

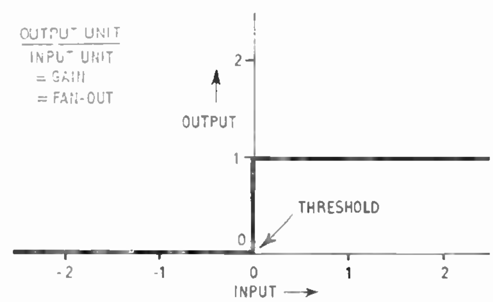


Fig. 1. Transfer characteristic of an ideal logic amplifier.

2. Attributes of Tunnel Diodes

The tunnel diode possesses two excellent conduction mechanisms which fail to overlap, thus giving a region of negative a.c. resistance which may be used to obtain amplification (Fig. 2). The signal amplitude of about $\frac{1}{2}$ V which may be obtained from a ger-

† Ferranti Ltd., Computer Department, Bracknell, Berkshire (now I.C.T., Bracknell).

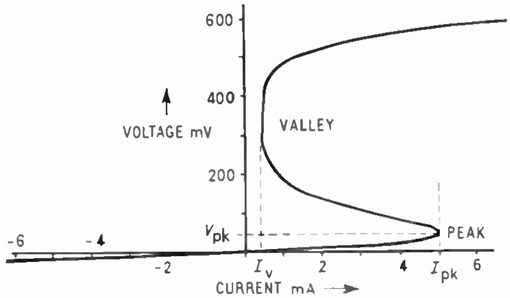


Fig. 2. Typical characteristic for a 5 mA tunnel diode.

manium tunnel diode corresponding to currents of a few milliamps in transmission lines with the impedance given above. This current level is suited to both tunnel diode and transistor technology. However, the two-terminal nature of the tunnel diode makes it unsuitable for matching to inputs of fixed impedance, and the same feature prevents it from acting as a directional amplifying device. It is important, when supplying these deficiencies, to avoid undue increase in the number of devices through which a signal must pass serially.

It might appear that the high-speed performance of tunnel diodes precludes the use of more familiar semiconductor devices in combination with them. A closer examination shows that switching times of less than 1 ns (under useful loading conditions) have only been obtained with diodes at a very expensive selection level. It may be assumed that improvements in manufacturing will make the 1 ns diode readily available in a year or two from now, but faster operation than this cannot be guaranteed. It is recognized that statements of this nature are habitually proved wrong by the semiconductor technologists, but in this case the stray capacitances of encapsulation and mounting are important for diodes which yield the correct signal currents for line matching, and this complicates the basic problem. The time scale so defined includes the latest types of switching transistor, the limiting time constants of which are smaller than 1 ns, thus making these devices suitable for directing current flow. Also, for signals of a few milliamps applied between base and emitter,

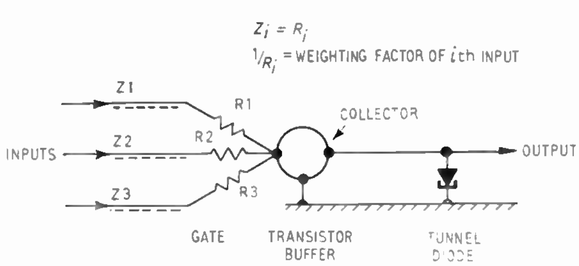


Fig. 3. Hybrid logic element.

such transistors present an input impedance of 5 to 15 ohms which approximates to a pure current sink for signals at over 100 ohms impedance. Input matching may therefore be achieved by terminating each signal line in a resistor of its characteristic value. A group of such resistors automatically forms a current-mix (or Kirchhoff) gate (Fig. 3).

3. The Transistor as a Signal Buffer

The arrangement of Fig. 3 shows that the introduction of a transistor can give excellent signal isolation at the expense of an additional transmission delay of less than 1 ns. It is natural to consider a grounded-base configuration for the transistor, since this leads to the minimum lag in passing input signals to the tunnel diode. However, the transistor is working with current signals and must therefore yield a gain of less than unity. It can be shown that the small, but non-zero, slope of the arms of the tunnel-diode characteristic lead to d.c. tolerance conditions which do not permit fan-out exceeding two or three. The logical function is also non-inverting and as such does not lead to a complete system in the absence of special measures.

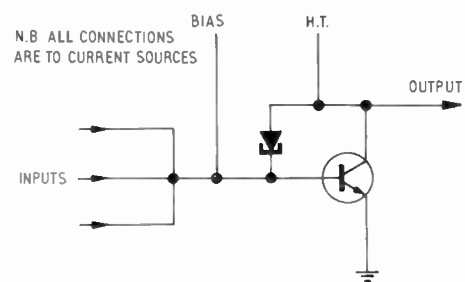


Fig. 4. Basic circuit of feedback amplifier.

If the emitter of the transistor is grounded there will be a somewhat longer delay in transferring a small signal to the collector owing to the reversed action of parasitic admittances. The delay can still be less than 1 ns if there is no voltage change at the collector (so that Miller feedback is negligible), and in the circuit under consideration this condition prevails until the tunnel diode switches. The operation is now an inverting one and the transistor is in a configuration admitting current gain, but the latter is likely to prove embarrassing; during fast switching Miller feedback could swamp the signal input to the transistor, while the settled-down output of the transistor would only increase the saturation current in the diode. The situation is greatly improved by returning the diode to the input node instead of directly to ground (Fig. 4). This arrangement places the gain mechanisms of the diode and transistor in series, while the important stray capacitances are placed in parallel. The only significant inductance in the circuit is that in the base-collector-diode loop, which can be made very small. As a result, one of the major disadvantages of tunnel diode circuitry—tendency towards self-oscillation—is largely eliminated. The input and output d.c. levels are separated by the tunnel diode voltage which is practically zero in one state, thereby eliminating the necessity for correcting networks. The collector is always at a positive potential relative to that of the base with the diode orientation of Fig. 4, and the transistor is able to work in a reasonably high-speed manner. Silicon epitaxial transistors can accommodate the diode in reversed orientation, but the collector voltage is in a slower region when the diode has switched. Although slower in returning from the high-diode-voltage state, this arrangement has a very compact voltage system with low power dissipation. Figure 4 shows an npn transistor as typical of the most satisfactory types which are readily available, but circuits for pnp transistors have equivalent properties, and both kinds have been constructed.

4. Functional Description of Circuit

The steady-state pattern of input and output currents in the circuit of Fig. 4 are illustrated in Fig. 5.† Currents arriving at the input pass almost entirely through the diode and transistor collector. Any current applied to the output terminal will again divide so that $I_{out} \times \alpha$ passes through the collector, in this case by-passing the diode. The manipulation of input and output standing currents can therefore effect any required balance between transistor and diode currents. The static relationship between current and both input and output voltages is shown in Fig. 6. With the circuit biased to point A the application of a negative-going input signal will initiate switching to point B. A step in the input voltage occurs as a result of the reduced transistor current, but this is only about 20 mV in amplitude.

Return from point B to point A can only follow a change of input current beyond I_V, and the circuit is seen to be bi-stable. This is advantageous in some applications, but not in others. If the diode switching current ($I_{pk} \rightarrow I_V$) is smaller than normal signal amplitude the circuit exhibits hysteresis rather than

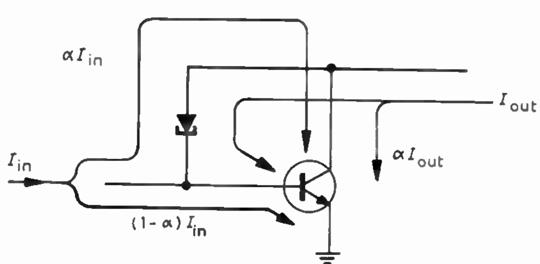


Fig. 5. D.c. input and output current distributions.

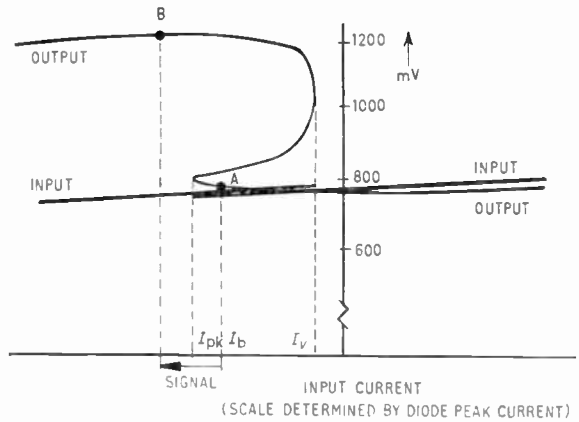


Fig. 6. Variation of input and output voltages with input current.

latching. The use of a backward diode will give an almost neutral characteristic (see Fig. 1), but no gain from the diode. This arrangement would logically result from an attempt to exploit the switching speed of the fastest available transistors by avoiding slow limiting mechanisms. W. R. Smith and A. V. Pohm‡ have adopted this approach, and report stage delays of as little as 3 ns with low fan-out, using resistor gates. They suggest that the use of higher current tunnel diodes would not contribute towards higher speeds, since an initial speed-up effect would be balanced by slowing down in the second half of the swing. This is true only if the diode switching current is small compared with the output signal current. When the two become comparable, the tunnel diode dominates the action in a manner which will now be described.

Initially, the transistor is required to raise the diode current to peak value at almost constant voltage. Any leakage of input current through the tunnel diode is in the correct sense, and the exact amount bypassing the transistor base is unimportant. As the diode peak point is approached, the current in the transistor must swing the output through several millivolts before the negative-slope region of the diode characteristic is fully reached. The diode then rejects current into the parallel combination of load impedance and input-output capacitance of the element. During the first part of the output swing the excess diode current will comfortably exceed the drain into the load, thus ensuring a rapid rise; this is a voltage-controlled operation which is independent of the time taken. Any signal current into the load

† See also Appendix 1.

‡ W. R. Smith and A. V. Pohm, "A new approach to resistor-transistor tunnel diode nanosecond logic", *Trans. Inst. Radio Engrs (Electronic Computers)*, EC-11 pp. 658–64, October 1962.

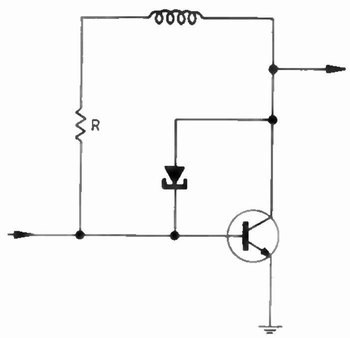


Fig. 7. Application of delayed negative feedback.

must return to ground via the base-emitter path of the transistor in the first instance, this current being additive to the input signal. This stimulates an amplified collector current which takes over as the diode valley region is traversed. In this way an almost linear rise is maintained for load currents exceeding that available from the tunnel diode alone. Following the observable signal swing there is continued adjustment towards the steady-state condition, but this occurs along the almost constant-voltage part of the characteristic of the tunnel diode.

If latching is undesirable, the inverting action of the circuit may be turned to good account by permitting negative feedback to be applied simply. An inductor is shown in series with a feedback resistor in the circuit of Fig. 7, the time-constant of the path ideally being about equal to the switching time of the inverter. Figure 8 shows the input current/output voltage curve, with a load line corresponding to the value of the feedback resistor. The final point of rest is now C instead of B, but the presence of the inductor ensures that the negative feedback does not have significant effect until the switching process is complete, speed being virtually unaffected. The inductor may, in practice, be a thin printed spiral or even a property of the resistor. Experimental evidence indicates that the inductor-like lag in response of the transistor may be adequate in itself. The application

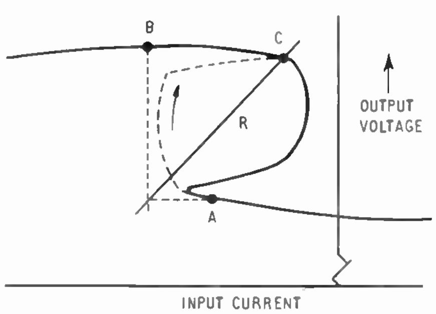


Fig. 8. Switching path with negative feedback.

of negative feedback is beneficial in that the tunnel diode is not permitted to conduct heavily in the high voltage state, which has a slower conduction mechanism than that at low voltages. The use of gallium arsenide diodes is also made possible by the avoidance of high forward currents, which would otherwise damage the junction after several months of operation.

The switching performance of elements using rather slow tunnel diodes has been predicted as approximately 6 ns.† Experimental evidence is confirmatory, although more extensive series of tests are required than have been possible to date. A practical limit of about 2 ns stage delay in the 'forward' direction (i.e. from low to high diode voltage) is anticipated, with the time for reverse switching dependent upon the negative feedback applied. Components for faster versions of the circuit than this are available, but costs are high and the transistors have base-spreading resistances of inconveniently large values. The voltage/time curve of the output signal (Fig. 9) is clean-cut owing to the regenerative action and the resistive nature of the load, and this assists in the prediction of timing margins.

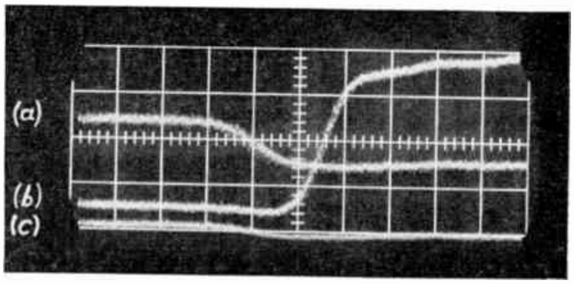


Fig. 9. Input and output waveforms (fast diode).

Triple exposure of sampling traces:
(*a*) Input signal applied through 47 Ω resistor.
(*b*) Output signal applied to 50 Ω cable.
(*c*) Input measured at transistor base.

Horizontal scale	1 ns/major division.
Vertical scale	Approx. 100 mV/major division.
Transistor	S.T.C. BSY 29.
Tunnel diode	R.C.A. 1N3128.

Time scale zero identical for all three traces.
D.c. level constant for traces (b) and (c).

Fan-out	3 (approx.).
10–80% output rise-time	1 ns (approx.).
Stage delay	1·6 ns between centre-line transits.

5. Methods of Improvement of D.C. Tolerance Limitations

The use of linear gates inevitably restricts fan-in as a result of d.c. uncertainties. By far the most important uncertainty is that of relative input voltages, which depends mainly on the spread of V_{be} and

† See Appendix 2.

partly upon the spread of current gain among the transistors. These uncertainties amount to a range of about 60 mV for silicon planar transistors, which is to be related to a signal swing of about 400 mV. A useful measure of the logical power of an element is given by the product of fan-in and fan-out, this figure being 4 with an input voltage range of 60 mV. The corresponding figures for 10 mV and 20 mV uncertainties are 30 and 15 respectively, and it will therefore be seen that measures for achieving closer tolerances should be very rewarding.

One method is to grade tunnel diode transistor combinations after assembly as part of the testing process to be applied to all completed elements. Each logic sub-assembly would be constructed of one grade only (the particular grade is immaterial) on the reasonable premise that connections will be logically simpler between units than within them. Another matching technique is to use pairs of transistors, as shown in Fig. 10. This arrangement gives a reference to ground at each stage, and manufacturers are already offering pairs of transistors in a single can, matched to 10 mV in V_{be}, for use in d.c. amplifiers, etc. The simple circuit of Fig. 10 may be made much more useful by taking a signal from the second transistor, the emitter current of which complements that of the first, thus giving approximately equal and opposite collector (output) current swings. Experiments indicate that the second output lags on the first by less than 1 ns, and possibly half that figure. Three circuits are shown in Fig. 11, the first of which is suitable for giving voltage amplification at the free collector; this may be of use in signal-level translation at electronic interfaces. The other two arrangements are for singly- and doubly-standardized outputs, the latter being suitable for majority gating decisions—a logical facility which is outside the tolerance limits with single unselected diodes of current manufacture.

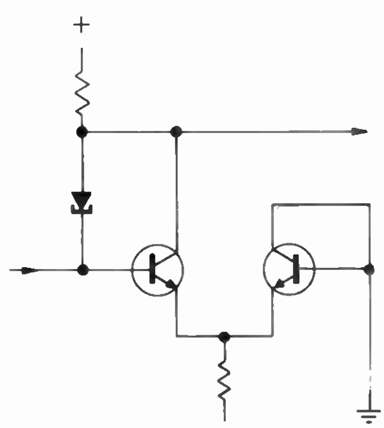
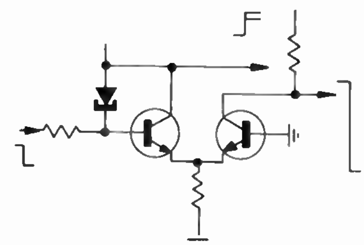
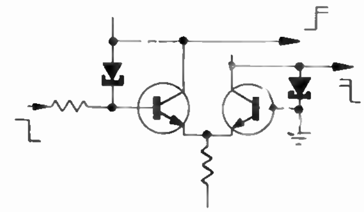


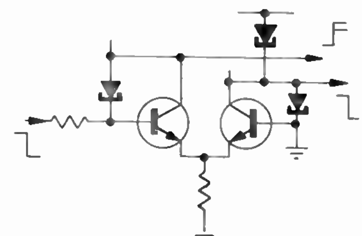
Fig. 10. Matched long-tailed pair.



(*a*) Voltage amplifier.



(*b*) Standard pair.



(*c*) Double standardization.

Fig. 11. Matched-pair circuits.

The speeded-up long-tailed pair, which has been seen to have evolved from voltage matching considerations, offers a variety of interesting logical configurations which would not be appropriate to discuss here. These include arrangements using both npn and pnp pairs, which follow easily from the common ground reference for both types. These circuits have a higher input impedance than the original grounded-emitter version, but adequate impedance matching is still possible.

Other measures for avoiding input voltage variation include the use of series combinations of diodes (or the use of high-energy-gap types) to increase the signal swing. Smith and Pohm have used this method successfully, but it is as expensive as the circuit of Fig. 10. An alternative for more specialized applications where large fan-in is required would be to employ multiple planar transistor gates with a compromise in matching to inputs. There is every reason to suppose that the high-speed amplification of the tunnel diode transistor circuit could be advantageously associated with current routing planar circuits of an uncritical nature.

6. Conclusion

In conclusion it may be stated that the non-linear amplifier formed by using a tunnel diode in the feedback path of a grounded-emitter transistor exhibits most of the properties required in ultra-high-speed logic circuitry. Initial experimental work indicates that the configuration is both stable and flexible, and that the speed obtainable should stand comparison with any other technique known today.

7. Acknowledgments

This paper describes the initial stages of a research project at the Bracknell Computer Laboratory of Ferranti Ltd., to whom the authors are indebted for permission to publish this paper. Much of the experimental work on the project has been carried out by Dr. F. E. Taylor of this Laboratory.

8. Appendix 1: D.C. Characteristics

The tunnel diode acts as a very-low-resistance strap between base and collector of the transistor, except when switching occurs. The relationship between base voltage and the algebraic sum of all currents flowing into the element shown in Fig. 4 is that of a transistor connected as a diode (Curve D, Fig. 12).

It is convenient to separate the currents applied directly to the base (inputs) and those applied to the collector (outputs), so that Curve D of Fig. 12 is broken down into a family of curves, A . . . A, each one of which corresponds to a fixed net output current. The running parameter of all these curves is input current, and the ordinates correspond to the input voltage of the element.

If h_{FE} is large, the input current flows almost entirely through the tunnel diode to the collector. The diode voltage may therefore be plotted to the same abscissae, as shown in Curve B. The output voltage of the element is the sum of input and diode voltages, the value of input voltage being selected from the appropriate A curve. In general two A curves will be involved, since the net output current is the sum of the standing current from h.t. and the output signal current, the latter being two-valued. The transition between the two A curves is directly reflected in the step in input voltage at the switching points as shown in Fig. 6.

The diode and base voltage curves have equal and opposite slopes near the bias point on both arms of the diode curve (see Fig. 6). This results in the output voltage curves being almost flat over the operating region, thus implying good isolation of output from input.

9. Appendix 2: Switching Time

The performance of the circuit was assessed for a combination of transistor with $f_T = 150$ Mc/s and a tunnel diode with $I_{pk} = 5$ mA; the sum of diode and transistor collector capacitances was taken to be 30 pF. The response of this arrangement to input current steps of between 1 mA and 3·3 mA for varying loads was calculated, and some of the results are given in Fig. 13.

The calculations were based on very simple assumptions, but experimental evidence (although limited at the time of writing) has confirmed the predictions of low sensitivity to power level and fan-out, and

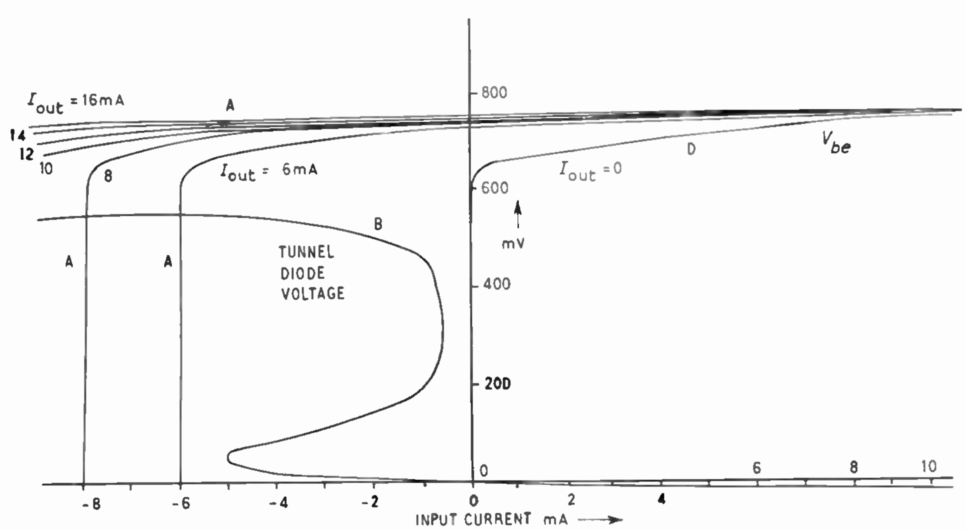


Fig. 12. Voltage/current relationships for transistor and tunnel diode.

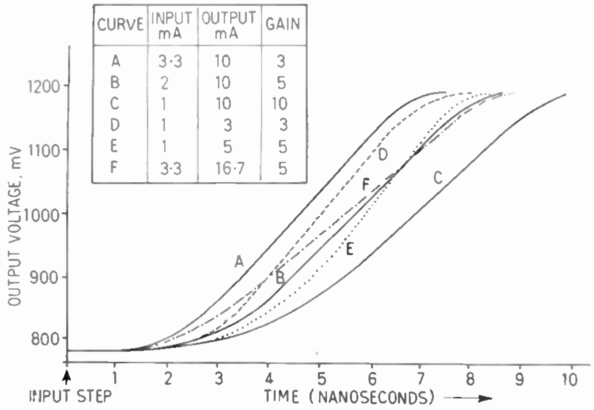


Fig. 13. Calculated response curves (slow diode).

actual switching delays appear to be within 30% of the calculated values in the worst cases. Figure 14 illustrates the method of calculation, the input and output currents being 1 mA and 3 mA respectively. The transistor tunnel diode combination is treated initially as a pair of *RC* circuits in series, the time-constant of the diode being taken as

$$R_S C = 6 \times 30 \times 10^{-12}$$
$$= 0{\cdot}18 \text{ ns}$$

where R_S = diode saturation resistance = 6 ohms,

C = total feedback capacitance = 30 pF.

The equivalent time-constant of the transistor is taken to be

$$\tfrac{1}{2}\pi f_T \simeq 1{\cdot}1 \text{ ns},$$

and $$\alpha \simeq 1.$$

The admittance of the load is negligible compared with that of the feedback path, and curve A is a plot of the current in the latter (i.e. through the diode). This current rises to half value in 1 ns, and the diode threshold is assumed to have been reached. The remaining current is available as 'overdrive', to which must be added the diode 'excess current' (curve B) which is obtained from the voltage/current characteristic of the diode.†

The calculations must now proceed numerically, since the required abscissa is time, step-by-step

† The published characteristic of the S.T.C. JK30A tunnel diode was used for this purpose.

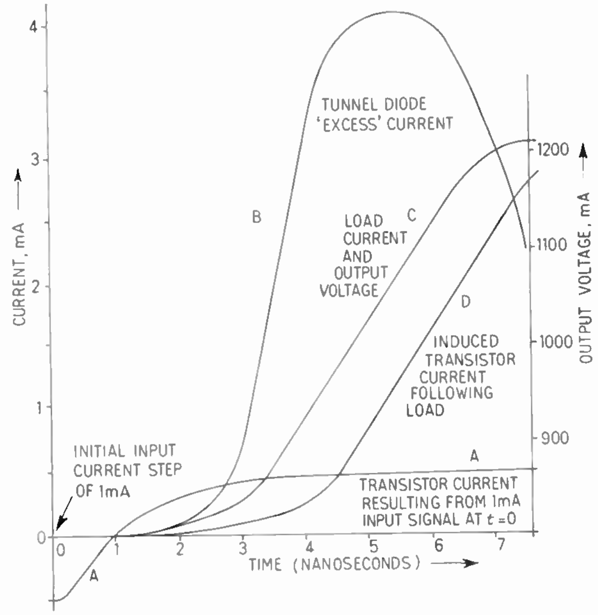


Fig. 14. Distribution of currents during switching.

estimates being made of the net current available for charging the 30 pF capacitance between input and output. Currents A and B contribute positively, but curve C represents the load current, which is subtractive. Curve C also represents the output voltage to the given scale. The contribution of the transistor due to the change in output current is shown in curve D (additive). Since all the load current must return to earth via the emitter of the transistor it is assumed that partition between base and collector occurs at a rate which, in an integrated process, can be represented by a delay equal to the transistor time-constant given above. D is therefore a copy of C drawn 1·1 ns later.

The numerical plots were made with ½ ns intervals, and one stage of back-correction was applied. Each plot required about one hour for completion, recourse to a computer being unnecessary.

Manuscript received by the Institution on 14th March 1963. (Paper No. 870/C62.)

U.D.C. 621.374:621.382.2:681.3

Some Properties of the Tunnel Diode Balanced Pair Circuit

By

N. E. WISEMAN,

M.S., B.Sc.(Eng.)†

Presented at a Computer Group Symposium on "Tunnel Diodes as Switching and Storage Elements" in London on 23rd January 1963.

Summary: The tunnel diode balanced pair circuit has been widely publicized as a potentially very fast computing element. On a very small scale it has been shown to work at more than 1 kMc/s. There are however several aspects of this circuit which make it difficult to use. This paper attempts to expose a few of the problems associated with the balanced pair as it might be used in an actual computing system. Some of these problems have been overcome, others are being investigated and it seems likely that a complete system based on tunnel diodes can be brought into operation in the reasonably near future.

1. Introduction

The tunnel diode balanced pair was proposed as a logic element over three years ago. Several papers have been published describing the basic operation of the circuit[1, 2, 3], and various tiny logic arrays have been built and made to work.

With suitable diodes, reliable operation at modest frequencies (say around 10 Mc/s) can be obtained with designs based on very simple theories of operation. However, diodes are available having gain-bandwidth products of several kMc/s and it might therefore be expected that the balanced pair circuit could be operated at very high speeds, say of the order of 1 kMc/s.

It is the purpose of this paper to draw attention to a few of the problems involved in the u.h.f. operation of the circuit. An understanding of the basic principles of operation will be assumed.

2. Stability

It has been shown[4, 5] that the inductive reactance which terminates a diode pair (i.e. across its pump terminals) must be small if self oscillations are to be avoided. An approximate value for the limiting inductance in the pump source is $L_s < R_s|R_n|C_b$ where R_s is the total series resistance, R_n is the diode negative resistance and C_b is the diode barrier capacitance. For even moderately fast diodes (5 mA, 10 pF) this gives $L_s < 2$ mμH which is almost impossible to attain at all package positions in a system of any size.

Furthermore, there is an additional difficulty arising from the connection of many circuits to the pump. During the transition of a circuit we may get undesired oscillations excited by the negative resistance of some of the diodes in the sundry reactances of the system. If, during the transition, a diode pair presents a negative resistance to its pump terminals, then there may be a limit to the number of circuits we can feed

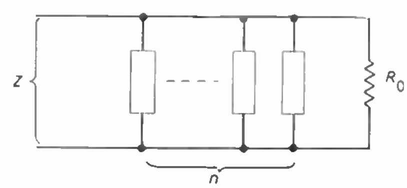


Fig. 1. Connection of a number n of tunnel-diode pairs to a pump transmission line.

from a single pump transmission line. Consider, for example, Fig. 1. Here n circuits are shown attached to a line terminated in R_0. The impedance seen at some point along the line is as shown in Fig. 2. Clearly if n is sufficiently large the net impedance of the line will be negative and may therefore excite oscillations in any stray reactances connected to it.

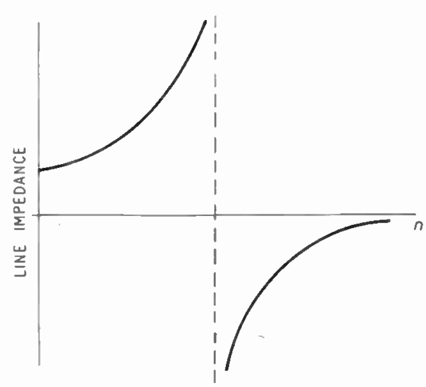


Fig. 2. Impedance at a point on the line in Fig. 1.

† University Mathematical Laboratory, Cambridge.

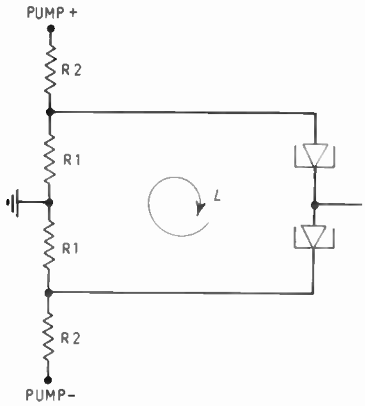


Fig. 3. Connection to the pump line to avoid instability.

The arrangement shown in Fig. 3 was adopted to avoid these difficulties. Here we have set $R_2 \gg R_1$ and made L as small as possible. The two important properties of this circuit are:

(*a*) over a wide range of impedances presented to the circuit by the pump system, the tunnel diodes are unconditionally stable.

(*b*) at all times during the transition of a pair the impedance seen looking into the circuit is positive and large (of the order of R_2).

Thus an arbitrarily large system can be constructed from these packages without instability problems (there are other problems as we shall see).

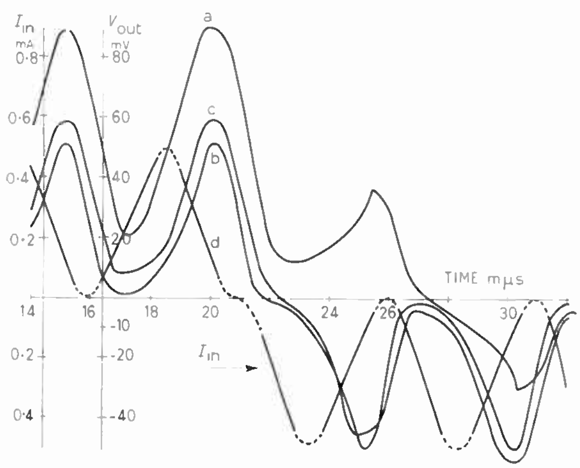


Fig. 4. Circuit switching characteristics

(*a*) Package adjusted for maximum gain.

(*b*) Package adjusted for maximum gain subject to zero storage effect.

(*c*) A compromise between (*a*) and (*b*).

(*d*) Input current curve.

3. Memory

A circuit with its pump in the ON state has one diode in the high state and the other in the low state. A circuit with its pump in the OFF state has both diodes low. Thus one of the diodes suffers a larger voltage change than the other during the pump transition ON → OFF. In even a perfectly balanced circuit the barrier capacitance of the two diodes will thus offset the voltage on the signal node (due to the unbalance in charge) as the pump turns off. As the operating frequency is increased this effect is aggravated and the circuit shows an increasing reluctance to change state from cycle to cycle. Usually the storage effect can be reduced at a given operating frequency by adjusting the pump parameters so as to reduce the signal swing—this gives a reduction in available gain. The behaviour of a package adjusted for maximum gain, for maximum gain subject to zero memory effect and for a compromise between the two is shown in Fig. 4.

4. Back Coupling

The normally proposed way of operating the balanced pair in a system is to use resistors as the coupling elements and a 3-phase pump to determine signal directivity. Resistors, being bilateral, couple circuits in both directions and we must watch out for sneak paths when determining a value for the coupling resistor.

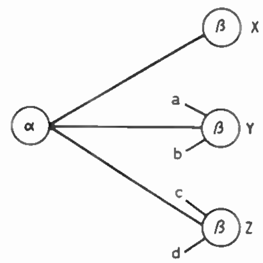


Fig. 5. Illustration of sneak paths.

Consider the arrangement shown in Fig. 5. Let us say the α-circuit is in the 1-state and that the β-pump is coming on. If the inputs a, b, c and d, are all zeros then circuits Y and Z should take up the 0-state. However, X should take up the 1-state. If the Y and Z circuits switch first then they interact with the α-signal and reduce its effect in switching X. This effect is known as 'back-writing' and can be very troublesome. An optimum value for the coupling resistor will depend on various circuit parameters and upon the back-writing effect and is quite difficult to compute.

5. Saddle Points

As either the frequency or the pump voltage is raised any given package will show an increasing tendency towards a sort of tristable operation where the circuit switches up, or down, or not at all, depending upon the sign and magnitude of the input current. The effect has to do with the switching time of the diodes and can best be explained by reference to the *V-I* characteristic at the signal node of the circuit at various instants during the pump transition. In Fig. 6 (*a*) the pump is low (off) and the circuit has only one stable output voltage (zero volts); in (*b*) the pump is at the critical value, when the circuit begins to switch; in (*c*) the pump is shown at some intermediate value, where, given sufficient time. the circuit would take up one of the two stable intersections A, A′ (0 being a saddle point). At a higher pump voltage, shown in (*d*), five intersections appear. If the circuit has switched beyond the saddle points B, B′, then one of the intersections A, A′ will ultimately be attained. If, however, the circuit has not switched beyond the saddle points, at the time they appear, then the circuit will fall back to intersection 0. The time at which the saddle points appear evidently decreases with increasing frequency, and also with increasing pump amplitude. It therefore follows that, if the diodes switch at a constant rate, the smallest pump amplitude at which tristable operation occurs must decrease with frequency. The general effect of this is to reduce the available gain for two reasons:

(*a*) the input current must be increased in order to raise the initial switching speed of the diodes,

(*b*) the pump voltage must be reduced below that which provides maximum output.

6. Present Status of Experimental Work

The arrangement shown in Fig. 3 was constructed with silicon dice resistors encapsulated in epoxy resin for R1 and R2. About 50 specimens have been individually tested at frequencies up to 500 Mc/s, and various rings and chains have been operated at up to 150 Mc/s. At 50 Mc/s arrays of about 30 circuits have been operated. With slightly faster diodes and improved tolerances reliable operation at 200–300 Mc/s is expected (in a system) and, with considerably improved diodes, operation at over 1 kMc/s is feasible. The system problems at 1 kMc/s and beyond are considerable and present experimental work seems to indicate that 200–300 Mc/s is a reasonable target.

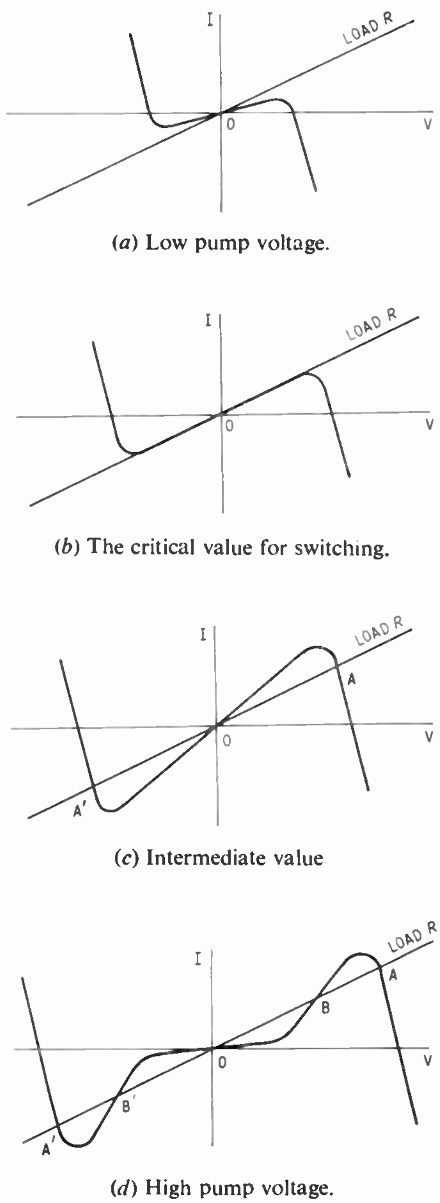


Fig. 6. Signal node characteristic at various instants during the pump transition.

7. References

1. E. Goto and others, "Esaki diode high speed logical circuits", *Trans. Inst. Radio Engrs*, **EC-9**, No. 1, p. 25, March 1960.
2. W. F. Chow, "Tunnel diode digital circuitry", *Trans. Inst. Radio Engrs*, **EC-9**, No. 3, p. 295, September 1960.
3. H. R. Kaupp and D. R. Crosby, "Calculated waveforms for tunnel diode locked pair", *Proc. Inst. Radio Engrs*, **49**, No. 1, p. 146, January 1961.
4. G. B. Herzog, "Tunnel diode balanced-pair switching analysis", *R.C.A. Review*, **23**, p. 187. June 1962.
5. N. E. Wiseman, "Some Experiments with Tunnel Diode Computing Circuits", Technical Memorandum 62/9. (University Mathematical Laboratory, Cambridge, 1962.)

Manuscript received by the Institution on 29th March, 1963. (Contribution No. 74/C63.)

U.D.C. 621.376.233

Carrier-Voltage Control for Improved Dynamic Range in Rectifier Modulators

By
D. P. HOWSON, M.Sc.,
(*Associate Member*)†

Summary: A technique for increasing the dynamic range of a balanced modulator by automatic adjustment of the carrier amplitude is described. A peak detector is used to sense the amplitude of input signals, and to control a variable-gain amplifier which reduces the carrier amplitude—and hence the carrier leak—for small inputs. It is shown that some commercially available modulator diodes are satisfactory for use in the system, although ideally bilinear diodes are required. Some experimental results on a phase-sensitive detector are included.

1. Introduction

In many applications of modulators using semiconductor elements it is desirable to minimize the percentage of the carrier voltage appearing in the output circuit. To achieve this aim, balanced modulator circuits are available, usually with potentiometer adjustments to obtain the best possible results.[1] Often this is sufficient for the application in mind, particularly as when the components of the carrier leak do not appear in the output frequency band they may be further reduced by filtration. However, there still remain important applications of balanced modulators where the balance thus obtained is not really sufficient, and many special circuits have been devised in an attempt to improve the situation.[2-7] Feedback from the modulator output to input via a second modulator has been used,[3] as have several modifications of the basic ring modulator circuit.[3-7] Other types of modulator element have also been used, as will be discussed later.[8-10]

This present paper is concerned with the possibility of reducing the carrier leak in modulators for small input signals, which is normally when it is most troublesome. In a normal modulator, the carrier voltage is determined by the maximum input signal that the device has to handle, and is therefore larger than is necessary for small input signals. Hence the leak is also larger than necessary for small inputs, and may be reduced by making the amplitude of the carrier vary according to the peak amplitude of the signal. This in turn may be achieved by the use of a device sensing the amplitude of the incoming signal, and using this information to modify the gain of an amplifier in the carrier path or, alternatively, if the signal amplitude varies in a predictable manner, as in an echo-ranging receiver, by varying the carrier voltage according to this law.[11, 12] In a modulator in which it is impracticable to separate the desired output product from one of the components of the carrier leak by means of a filter, therefore, reduction of the leak for small signals will increase the dynamic range of the device, where dynamic range is defined as

$$20 \log \left\{ \frac{\text{maximum amplitude of wanted output product}}{\text{minimum amplitude of carrier leak component at wanted product frequency}} \right\} \quad (1)$$

This definition assumes that system performance is unsatisfactory when the leak component is greater than the wanted output product, and that the minimum wanted product amplitude is not limited by noise considerations.

A considerable saving of carrier power will be made in any system in which large input signals are infrequent, by the use of carrier voltage control.

2. Carrier Voltage Control (C.V.C.) with Ideal Bilinear Rectifiers

The function of the carrier voltage in a rectifier modulator is to switch the diodes from a low value of incremental or a.c. resistance to a high value, and vice versa. The efficiency of the device as a modulator is dependent upon the ratio of these incremental resistances, which should be as high as possible. Considering now the proposed system of reducing carrier leak, which involves varying the amplitude of the carrier voltage, it is clear that it is desirable

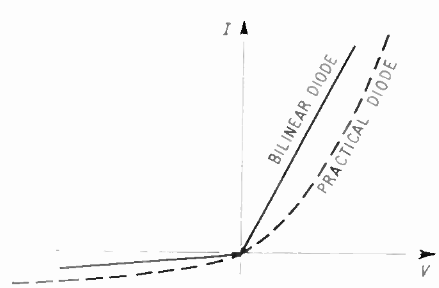


Fig. 1. Diode characteristics.

† Electrical Engineering Department, University of Birmingham.

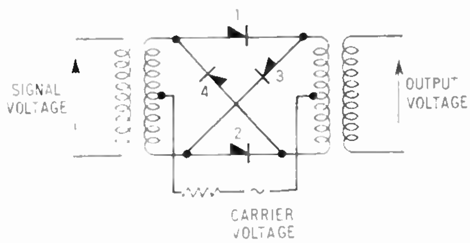


Fig. 2. Ring modulator.

that the efficiency of the device remain constant over a range of values of this voltage. This implies that the ratio of incremental resistances remains constant over the range, and in particular that the ratio remains constant for very small carrier voltages, to achieve the maximum dynamic range. A mathematical model of a rectifier which meets these requirements is the well-known 'bilinear rectifier' whose characteristics are depicted in Fig. 1. No real diode, of course, will have the desirable discontinuity of the model at the origin. If the ring modulator of Fig. 2 is considered, with a square-wave variable-amplitude carrier source, then the leak voltage waveform will also be a square-wave, assuming that the diodes, although all bilinear, differ in their forward and reverse resistances. (It is desirable that the carrier signal be a square wave to ensure that the signal voltage across each diode is smaller than the corresponding carrier voltage over the whole of each cycle. If it is not, distortion will take place, the diode switching being no longer wholly controlled by the carrier signal.) As the carrier amplitude V_C changes, the amplitudes of the leak voltage in each half cycle will change and will be directly proportional to V_C. The d.c., fundamental and higher harmonic components of the leak waveform will also, therefore, be directly proportional to V_C. The efficiency or conversion loss of the modulator will, of course, be independent of carrier amplitude since the incremental resistances on which these properties depend are similarly independent. (These results will also be true for shunt or series modulators similarly used.) For normal use when output and leak are not at harmonically related frequencies the modulator output will remain a linear function of the input, but will become a bilinear function if the device is being used as a phase-sensitive rectifier, as shown in Fig. 3. Since the maximum leak voltage is usually a very small fraction of the maximum signal voltage, this departure from linearity will be very slight and has been over-emphasized in the figure in order to show it at all.

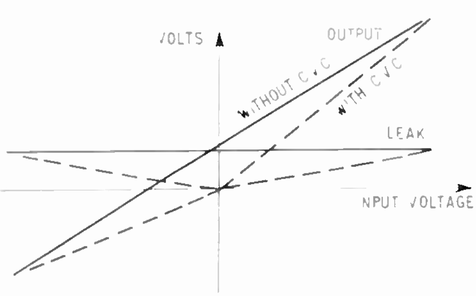


Fig. 3. Phase-sensitive detector characteristic using bilinear diodes.

3. Modification of the Results for Actual Diode Characteristics

The I-V characteristic of a typical semiconductor diode is shown in Fig. 1, and it is apparent from this that the incremental forward resistance of the diode increases and the reverse resistance decreases, as the carrier voltage tends to zero. Thus, depending on the shape of the diode characteristic and the type of modulator circuit used, there will be a lower limit to the carrier voltage below which it will not be practicable to go, due to the increase in conversion loss.

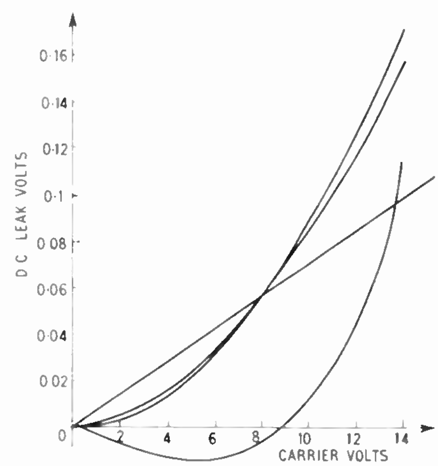


Fig. 4. Variation of d.c. leak with carrier voltage foı four different combinations of a set of diodes.

The components of the leak voltage will no longer be directly proportional to the carrier voltage as when bilinear rectifiers were postulated, and it is possible, though unlikely, that for a small range of carrier voltage a decrease in this voltage will lead to an increase in leak. Figure 4 illustrates the d.c. leak component as a function of carrier voltage for a ring modulator with four different combinations of diodes, and it will be seen that the point just made does occur in one case. Calculation of the leak depends upon the form of carrier voltage and the load impedance—for simplicity it will be assumed here that a square-wave carrier generator is used. Then the carrier leak waveform will still be a square-wave, but the amplitude will no longer be directly proportional to V_C. Using the leak formula derived in the Appendix, and assuming that the forward resistance of the diodes is much less than the modulator terminating resistance, during one half-cycle of carrier

$$(V_{CL})_1 \simeq \frac{r_2-r_1}{r_1+r_2} . V_C \qquad \ldots\ldots(2)$$

and during the other

$$(V_{CL})_2 \simeq \frac{r_4-r_3}{r_3+r_4} . V_C \qquad \ldots\ldots(3)$$

(where r_1, r_2, etc., denote the forward resistance of diodes 1, 2, etc., in Fig. 2).

Now for a batch of germanium point-contact diodes it has been shown[13] that the forward resistance may be represented by

$$r_m = R + a_m e^{-bV_C} \qquad \ldots\ldots(4)$$

where R and b are relatively constant for all the diodes of the same type, and a_m varies widely from diode to diode.

Then

$$(V_{CL})_1 \simeq \frac{(a_2-a_1)V_C}{(a_1+a_2)+2R\,e^{bV_C}} \qquad \ldots\ldots(5)$$

and $(V_{CL})_2$ will be of similar form. Since R is approximately the forward resistance of the diodes at 1 volt bias, a_m is the resistance of one of the diodes with zero bias, and b is a parameter which need only be evaluated once for a particular type of rectifier, the amplitude of the leak waveform in each half cycle may now readily be calculated. Since the leak waveform is a square wave, the d.c. component is

$$(V_{CL})_{\text{d.c.}} = \tfrac{1}{2}\{(V_{CL})_1 + (V_{CL})_2\} \qquad \ldots\ldots(6)$$

care being needed to give the correct signs to $(V_{CL})_1$ and $(V_{CL})_2$. Similarly, the amplitudes of the a.c. components are given by

$$(V_{CL})_{\text{a.c.}} = \frac{4}{t\pi}\{(V_{CL})_1 - (V_{CL})_2\} \qquad \ldots\ldots(7)$$

where t, which is odd, is the order of the harmonic of the fundamental leak waveform that is required. There are no even harmonics present according to this idealized model of the modulator.

The range of carrier voltage that may be used in a particular modulator also requires calculation. This will depend on the range of signal voltage that it is desired to work over, which could be at least 100 to 1. The desired maximum and minimum carrier voltage across each rectifier can be determined from this, bearing in mind the maximum allowable dissipation given in the manufacturer's data for the diode. It may not be possible to allow the carrier voltage to fall as low as would be desired, due to the worsening of the ratio (n^2) of the incremental forward and reverse resistances of the diode that occurs as the carrier voltage is reduced. This ratio will affect the conversion loss of the modulator[1, 14]—for instance for a ring modulator terminated with constant resistances the minimum loss is

$$3{\cdot}92 + \frac{17{\cdot}4}{n}\ \text{dB} \qquad \ldots\ldots(8)$$

In many applications it will be necessary to keep n large enough—say greater than 10—so that the conversion loss remains sensibly constant. The considerations outlined above affect the choice of diodes for a c.v.c. modulator. The silicon junction diode, for instance, is high resistance until the carrier voltage exceeds about +0·6 volts, and so is not very suitable. Germanium point-contact diodes are quite satisfactory for carrier voltage ratios of some 15 : 1, but the incremental resistances change appreciably over this range due to the curvature of the diode characteristic. Germanium junction diodes have higher values of n, but due to their low forward resistance cannot be used with large voltages across them. If however a suitable series resistor is added, the combination is a good approximation to the desired bilinear rectifier. For example, a ring modulator using Mullard AAZ 12 germanium junction diodes in series with 150 Ω resistors may be used with a 100 : 1 carrier voltage ratio—and hence at least a 100 : 1 signal voltage ratio. The calculations necessary to establish this go as follows. If the resistors are within 1% of each other, selecting the worst combination, from eqn. (7) the fundamental component of leak voltage will be 37·9 dB below the peak carrier voltage. If the maximum signal voltage is 10 dB below the peak carrier voltage therefore, the leak component is 27·9 dB below the maximum, and 12·1 dB above the minimum signal voltage, if the modulator is used without c.v.c. With c.v.c., the situation for the maximum signal voltage is unchanged. For the minimum signal voltage on the other hand, if the carrier voltage is reduced by a factor of 100, the forward resistances of the diodes ≃ 150 Ω. If the diode resistances are within 10% of each other for the reduced carrier voltage then, by selecting the worst possible combination, the fundamental component of leak, from eqn. (7), will be about 13 dB below the smallest signal voltage, without any balancing controls having had to be introduced into the circuit. The conversion loss of the modulator, if it is of the constant resistance type, should vary by less than 0·25 dB over this range of carrier voltage,[1] so that the wanted output product voltage should be in all cases some 4 dB below the signal voltage, from eqn. (8).

In an ordinary ring modulator it is difficult to obtain a large negative voltage across the back-biased diodes due to the forward biased diodes inevitably in parallel. This restriction does not occur with the series or shunt modulators where a large negative voltage may be obtained by the use of a suitable value of carrier source resistance, without passing too much current through the diodes when forward biased. It is worth pointing out here that the modified ring modulator circuits mentioned previously, and discussed in more detail elsewhere,[1] are similar to the latter modulators in this respect, and for this reason alone their use may be preferable in double-balanced modulator circuits using carrier voltage control, since in this way a greater ratio of forward to reverse incremental resistance may be obtained.

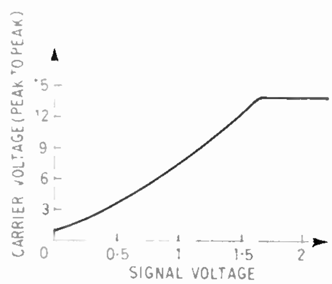


Fig. 6. Characteristic of the variable-gain amplifier.

4. Experimental Results for a Phase-sensitive Detector

Carrier-voltage control appears especially suited for use with low frequency modulators, of which phase-sensitive detectors form an important group. Accordingly, a phase-sensitive detector with carrier voltage control was built and tested as an undergraduate project by A. F. Newell of this Department. The block schematic of the system is given in Fig. 5, the phase-sensitive detector being a ring modulator with constant resistance terminations using OA81 germanium point-contact diodes. The carrier frequency used was 1 kc/s. The carrier amplifier was

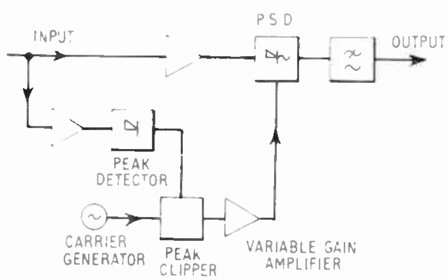


Fig. 5. Phase-sensitive detector with carrier voltage control.

required to have a gain varying linearly with respect to a bias signal, which in turn was obtained from a peak detector operating on the signal input.

The amplifier consisted of a single pentode stage, fed from the carrier generator by a large voltage, the positive half-cycles of which were clipped, the flat top thus created being clamped to the level of the sum of the voltage produced by the peak detector and a negative d.c. bias. The amplifier therefore produced a square-wave output with an amplitude approximately proportional to the difference between the cut-off voltage and the clamping voltage, i.e. approximately proportional to the incoming signal amplitude. For large inputs the pentode over-loaded, preventing damage to the diodes: for small inputs the valve operated on the curved part of its characteristic so that a carrier voltage was applied to the diodes which was larger than would have been obtained if the amplifier variation had been truly linear. This was necessary, as has been discussed earlier, to keep the conversion loss of the detector from increasing unduly. Figure 6 gives the final amplifier characteristic. (Valves were used in the circuit for reasons of economy only.)

The peak detector was conventional and was preceded by a buffer stage to avoid loading the signal circuit. The buffer amplifier in the signal path was inserted to avoid oscillatory feedback around the loop evident in the block schematic, which occurred otherwise at signal frequency, carrier leak occurring in the input transformer of the phase-sensitive detector. The transient response of the peak detector circuit and the variable gain amplifier was such that the total rise-time of the two circuits was 70 ms and the decay time 250 ms.

The phase-sensitive detector operated with a maximum signal of 2·4 volts and with 7 volts peak to peak carrier voltage across each diode. Referring to Fig. 2, to allow a d.c. output the signal and switching voltages were fed in via the transformers and the output was taken across the centre-tappings. The zero signal output voltage of the original device—the d.c. carrier leak—was equal to 180 mV, whereas when carrier voltage control was added this fell to 1·4 mV. This is therefore a 42 dB improvement in the dynamic range of the device. Careful matching of the diodes initially could have improved the absolute leak figures, but little change would be expected in the magnitude of the improvement achieved. The characteristic of the detector with and without carrier voltage control is shown in Fig. 7—the non-linearity evident would, of course, have been

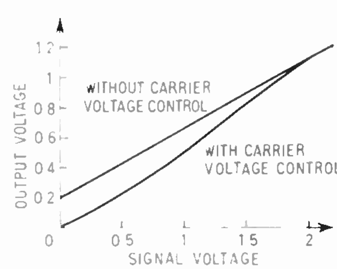


Fig. 7. Improved phase-sensitive detector characteristic.

much less with selected diodes. Note that the percentage amplitude distortion of the signal is now, with c.v.c., almost independent of the signal amplitude, and does not increase as the signal decreases as happened in the original detector. The d.c. leak component for a maximum-amplitude quadrature signal is, of course, not improved by carrier voltage control.

5. Experimental Results for a Single-sideband Modulator

During the development of modulators for an echo-ranging system, Chatterjee[12] improved the dynamic range of a transformerless shunt modulator[1] using a technique similar to that described previously. Here it was desired to produce a single-sideband modulator with a wide dynamic range, using a swept local oscillator over such a frequency range that no filtration of the fundamental component of the carrier leak was possible (see Fig. 8). Tucker *et al.*[11] had suggested that as the input signals to the modulator decreased in strength in a regular fashion over the period of a sweep it should be possible to improve

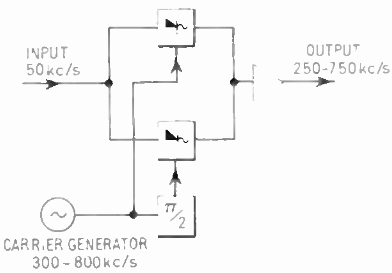


Fig. 8. Single-sideband modulator.[12]

the dynamic range of the modulator by decreasing the carrier amplitude in a similar manner. The results obtained were rather disappointing, as although at 300 kc/s the dynamic range was improved from 46 to 58 dB, at 800 kc/s it was only improved from 28 to 36 dB. It seems probable that capacitive effects figure largely in the poor results obtained at 800 kc/s.

6. Comparison with Other Methods of Improving Dynamic Range

A recent paper[3] has proposed a feedback system for increasing the dynamic range of a phase-sensitive detector (Fig. 9). In this the output from the phase-sensitive detector is algebraically added to the signal input, and for one half-cycle of the carrier the resultant is fed back to the signal input. During the other half cycle a shunt modulator in the feedback path shorts out the feedback. Thus during the half cycle in which the input and output of the modulator used

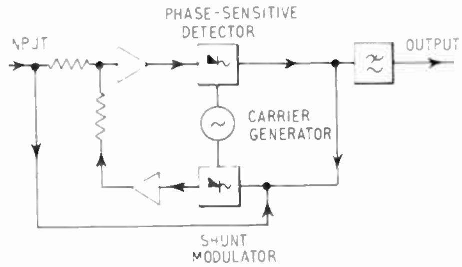


Fig. 9. Jones' feedback scheme.[3]

as the phase-sensitive detector are in opposition, their difference, which represents the modulator leak, is minimized by negative feedback. During the other half cycle in which they would add the feedback is inoperative. This technique is claimed to reduce the null voltage by a large factor and there is no doubt that it will do so over the one half-cycle, since the shunt modulator in the feedback path, which need only handle small signals, will have a correspondingly smaller null error itself. However, in averaging the null error over a full cycle, it is clear that its d.c. component can at best be reduced by one half. The addition of a second similar feedback path operative for the second half-cycle is possible and would theoretically produce the desired improvement in dynamic range but the complication of the system is increased and the possibility of spurious oscillation also.

The technique of feeding the diodes in a ring modulator from two identical constant-current generators in order to reduce the carrier leak has been suggested,[1, 2] and there is no doubt that good results are obtainable by these means. For applications where a rapid transient response is essential this system is probably superior, but for other work it is felt that control of carrier voltage is simpler and the circuit design and construction less critical. For example, the use of transistors for the auxiliary circuits for c.v.c. seems straightforward, which is not the case for the constant-current generators due to the high output impedances required.

The use of transistors, particularly symmetrical transistors, as modulating elements[8] has resulted in balanced modulators with very low leaks, as has the use of backward diodes.[9, 10] Both of these elements require only some 50 μW of carrier power, and although the backward diode can only handle a correspondingly smaller signal, the fact that it is nearly insensitive to temperature changes allows an excellent balance to be achieved by potentiometer adjustments alone. Further improvement in the dynamic range of a modulator using transistors could, of course, be achieved by carrier voltage control.

7. Conclusions

It has been shown that by carrier voltage control the dynamic range of a rectifier or similar modulator can be significantly improved. Signal amplitude distortion has been shown to be lessened with carrier voltage control, but harmonic distortion may occur, depending on the transient response of the peak detector used. The construction of the circuits required for sensing the signal amplitude and varying the carrier voltage have been shown to be simple and non-critical, which has not been the case with some other systems proposed to achieve the same ends. It has been shown that the theory of the system is simple when the rectifiers are assumed bilinear, but that the difficulty of computation increases rapidly when more realistic assumptions are made. Nevertheless, from the calculations that have been made, it is clear that a careful choice of diode or diode-resistance combinations will be needed to achieve optimum performance. Difficulties which may arise at frequencies where rectifier capacitance is important have not yet been investigated, and other work which has been cited suggests the effects may be considerable. There seems no reason however why the system should not operate satisfactorily with carrier frequencies of up to tens of kilocycles.

8. Acknowledgments

The author gratefully acknowledges the help received from Professor D. G. Tucker, whose comments and criticisms of this work in its early form were most useful. The author is also indebted to Mr. A. F. Newell for the experimental work quoted in Section 4.

9. References

1. D. G. Tucker, "Modulators and Frequency Changers", (Macdonald, London, 1953).
2. E. Hands, "A rectifier modulator with stable low carrier leak", *Proc. Instn Elect. Engrs*, **105C**, p. 381, September 1958.
3. S. Jones, "On the elimination of null in modulating and demodulating devices", *Trans. Amer. Inst. Elect. Engrs*, **80**, Part 1, p. 135, May 1961.
4. Philips Lamps Ltd., British Patent No. 622,711 (1949).
5. E. S. Grimes, U.S. Patent No. 2,462,093 (1949).
6. Standard Telephones & Cables Ltd., British Patent No. 471,548 (1937).
7. B. B. Jacobsen. Discussion on papers by D. G. Tucker, *Proc. Instn Elect. Engrs*, **96**, Part III, p. 218, 1949.
8. J. F. O. Evans, D. A. Gill and B. R. Moffitt, "Symmetrical transistors as a.c. or d.c. switches and their applications in modulator and demodulator circuits", *J.Brit.I.R.E.*, **21**, No. 2, p. 143, February 1961.
9. "Silicon uni-tunnel diodes for low-level bridge modulation", Application Note by Hoffman Electronics Corporation.
10. D. P. Howson, "Backward-diode modulators", *Industrial Electronics*, **1**, p. 334, 1963.
11. D. G. Tucker, V. G. Welsby and R. Kendall, "Electronic sector scanning", *J.Brit.I.R.E.*, **18**, p. 469, August 1958.
12. A. Chatterjee, "An investigation of problems relating to the design of acoustic scanning systems", M.Sc. Thesis, University of Birmingham, 1959.
13. D. G. Tucker and G. F. Machen, "The constancy of small rectifiers", *J. Sci. Instrum.*, **25**, pp. 369–71, 1948.
14. D. P. Howson, "Conversion loss variation in a rectifier modulator", *Industrial Electronics*, **1**, pp. 46 and 101, 1962.

10. Appendix

Considering the equivalent circuit for the ring modulator given in Fig. 10, we shall neglect the effect of the reverse-biased diodes, and the shunting effect of the transformers.

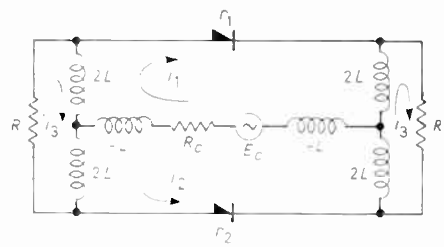


Fig. 10. Ring modulator equivalent circuit.

By symmetry the currents in the end loops are the same. Assume the transformers have negligible shunt reactance. Then from one end loop,

$$i_3 \simeq \tfrac{1}{2}(i_1 - i_2) \qquad \ldots\ldots(9)$$

Going round the outermost loop

$$2i_3R + i_1r_1 - i_2r_2 = 0$$

$$(i_1 - i_2)R + i_1r_1 - i_2r_2 = 0 \qquad \ldots\ldots(10)$$

Round the two inner loops, where V_C is the voltage across the rectifiers,

$$2V_C = i_1r_1 + i_2r_2 \qquad \ldots\ldots(11)$$

From (10) $i_1(R+r_1) = i_2(R+r_2)$ ......(12)

$$2V_C = i_1\left\{r_1 + r_2 \cdot \frac{R+r_1}{R+r_2}\right\}$$

$$i_1 = \frac{2V_C(R+r_2)}{r_1(R+r_2)+r_2(R+r_1)}$$

$$= \frac{2V_C(R+r_2)}{R(r_1+r_2)+2r_1r_2} \qquad \ldots\ldots(13)$$

Therefore

$$i_3 = \frac{V_C(r_2-r_1)}{R(r_1+r_2)+2r_1r_2}$$

and

$$V_{CL} = i_3R = \frac{V_C(r_2-r_1)R}{R(r_1+r_2)+2r_1r_2} \qquad \ldots(14)$$

Manuscript first received by the Institution on 23rd May 1963 and in final form on 12th August 1963. (Paper No. 871.)

U.D.C. 621.7: 681.3

Data Processing for Numerical Control of Machine Tools

By
D. F. WALKER, B.Sc.†

Based on a paper presented at the Symposium on "Recent Developments in Industrial Electronics" in London on 2nd–4th April 1962.

Summary: Numerical control is another form of automation which is most applicable to small batch productions. The paper describes the data processing for a typical control system and the design requirements on which it is based. General-purpose computers are essential for the process but are not in themselves sufficient, and special purpose curve generators are used to generate the continuous information for the control of the machine tool. The incremental computing technique employed is briefly described. The data processing costs are by far the largest item when the batch number is small and an indication is given how future development will reduce these costs.

1. Introduction

The use of numerical methods for controlling machine tools is aimed at making the manufacturing procedure more automatic and removing the human skill from the process. Initially, automation was carried out extensively by mechanical means to remove the human element. Numerical methods now offer a new opportunity to automate small batch production where previous methods were unsuitable because of economic reasons.

It is essential that any numerical control system should be designed to require the minimum amount of human intelligence and effort, and it should compare favourably in this respect with more conventional techniques. For instance, it is of little value to replace the skilled machine operator by an equally, if not more, intelligent planner who specifies the information for numerically controlled machine tools. In this context, the ease and simplicity of the planning system cannot be too highly stressed. This is one of the main reasons why computers are necessary for data reduction if full benefit is to be obtained from numerical methods. This has been one of the main design concepts behind the system to be described.

Machine tools may be divided broadly into four classes; lathes, drilling machines, milling machines, and transfer and special machines. Milling machines lend themselves to numerical control techniques and the system has been applied to a wide variety of milling applications ranging from simple components to three-dimensional turbine blade dies. Mechanical automation has been applied much more widely to the other classes of machines and the gain in these cases is not so favourable. The cutting of plates in the shipbuilding industry is an example of a special application of this technique. Special attention has been paid to the integration of the systems with present manufacturing techniques and a numerically controlled flame-cutting machine has been designed specially to meet the application.

† Ferranti Ltd., Numerical Control Department, Thornybank, Dalkeith, Scotland.

2. General Design Requirements

The following factors are considered essential for satisfactory operation of a numerical control system.

(1) The controlled machine should machine at least to the same accuracy as, if not better than, the conventional machine it replaces and should be capable of operating at higher feed rates.

(2) As well as controlling the three axes of the machine, a small number of miscellaneous functions, such as spindle on/off, coolant on/off, should be catered for.

(3) The controlled machine and data processing system should be matched to each other.

(4) The input information should be specified as easily as possible directly from drawings and with the minimum amount of manual effort.

(5) Human errors are the major source of trouble in any production system and as much as possible of the human element should be removed from the system. Checking and redundancy should be built into the system to take care of the human element left in the processing.

(6) Failure due to electronic causes should always be overcome by design so that they are self-correcting either automatically or manually.

(7) Equipment on the shop floor should be as simple as possible and capable of maintenance by the machine operator.

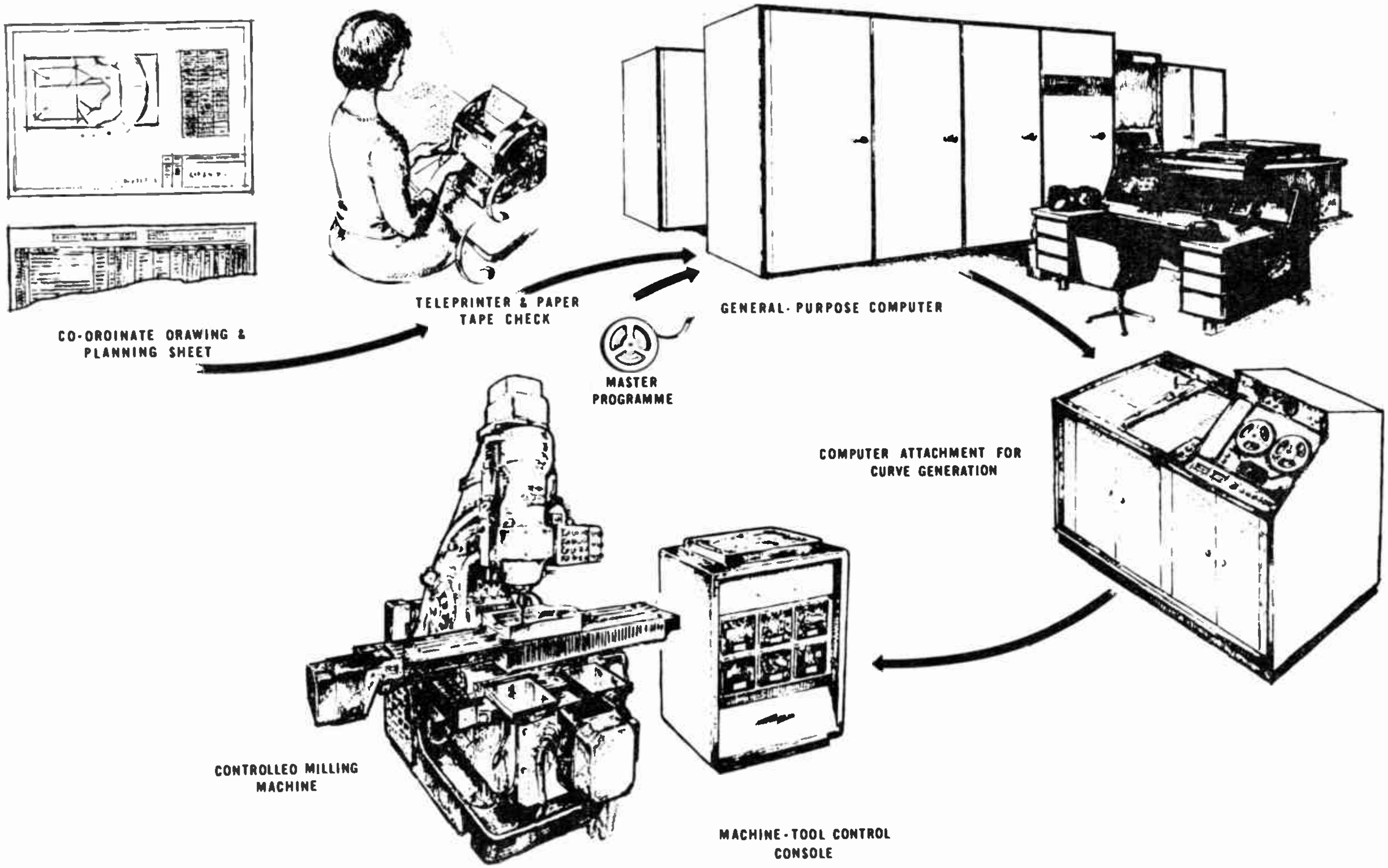


Fig. 1. Functional diagram of numerical control system.

(8) Whatever complicated equipment ere this should be at the data processing centre where a qualified and expert staff is available for its maintenance.

An accuracy requirement of at least 1 in 10^5 demands that digital techniques should be used. However, digital methods are complicated and should be avoided in preference to the much simpler analogue circuitry on the shop floor. A phase-modulated control system which would give the required accuracy for control was adopted for this reason.

The general-purpose computer is essential to keep the data processing simple, easily understood and enable the simple planning information to be reduced into a form suitable for controlling the machine tool. However, general-purpose computers are not capable of doing the complete data processing as they are very inefficient at generating the continuous information required for controlling the machine tool. After the tool off-sets have been calculated and the profile to be cut reduced to a series of equation constants, a special digital differential analyser or curve generator generates the continuous information for controlling the machine tool. The curve generator is basically a real-time computer which is very similar to its analogue counterpart in which the analogue integrators are replaced by digital ones. This type of computer is not restricted in its ultimate accuracy and in the case of the machine tool application would be set up to solve the differential equations, the constants of which have been specified by a general-purpose computer.

The phase system requires a large amount of information to be specified at the machine tool. Magnetic tape, which is the most suitable storage medium for this purpose was adopted. Although the information recorded on magnetic tape is highly redundant this is well worth the sacrifice to keep the controls at the machine simple. It also allows time-sharing of the computer among a large number of machines by using a much higher record speed than the replay speed at the machine tool. Special machine function signals can be recorded on the magnetic tape without interfering with the control information. As the digital elements of the computer can operate very much faster than required by the machine tool controls, time-sharing was employed to get the best utilization of the equipment. The simplest and most expedient way of doing this was to time-share the general-purpose computer and digital differential analyser type curve generator among a large number of control equipments using a magnetic tape link to couple the control units to the d.d.a.

The capital cost of such a data processing system is very large and can only be warranted for servicing a large number of machine tools. To allow for customers who have only one or two machines, a data processing centre which customers can use for processing their tapes was set up. The punched paper tape of the planning information is sent to the centre, processed through the general-purpose computer and digital differential analyser and a magnetic tape recorded. This tape is then returned to the customer for playing on the control console. It should be noted that the customer has complete control over the planning, and the data processing centre only transforms the planning information into pulses recorded on the magnetic tape without any planning interference from the centre, whose main purpose is to maintain the equipment.

3. Description of Numerical Control System

The functional diagram of the system is shown in Fig. 1. From the drawing of the component to be cut the planner fills in a planning sheet. This describes the geometry of the component and the method in which it is to be cut. A data tape is then punched from this planning information and fed into the computer along with a master program tape. The program tape writes into the computer the necessary programs for interpreting the data tape and carrying out all the necessary calculations and reduces the information into suitable form for feeding to the curve generator.

The choice of the link between the general-purpose computer and the curve generator depends very much on the application and the type of computers used. It can be magnetic tape, paper tape or a direct electrical link. The curve generator calculates the necessary continuous control signals and records them on the magnetic tape. At the same time a drawing is produced of the cutter centre path. This is returned to the planner and is a very important check as it allows him to see at a glance whether he has made any serious mistakes in his program. The recorded tape is then replayed on the control console of the machine tool and the component cut.

There are four information channels on the magnetic tape,[1] one control channel for each axis of the machine and a common reference channel. The signals from the reference channel generate a two-phase 100 c/s supply for feeding to the synchronous motors of the optical measuring units. A rotating disc with a spiral grating scans the linear reflecting grating to generate phase-modulated feedback-control signals, which are compared with similar signals from the appropriate control channel of the magnetic tape. Any phase error between the signals generates an analogue error signal for feeding to the hydraulic valves controlling the table servos. Thus the servo error signal corrects for any error between the feedback and demanded control signal and the machine table moves in unison with the control signals on the magnetic tape.

4. Data Processing Programs

The general layout of data processing subroutines is shown in Fig. 2. These are built up of basic programs which are common to many applications. The basic subroutines are self-contained so that they can be assembled in a manner depending on the application. Thus the output routine to punch the information for the curve generator is self-contained and can be added to any of the processing subroutines. They are assembled to flow into each other and are integrated into the particular manufacturing process by further specially written input programs. In the case of the Ferranti standard planning, the general-purpose computer interprets the planning information, calculates the change points between the mathematical curves making up the profile, calculates the off-set path to take care of the cutter diameter and specifies the constants of the differential equations of the various curves which make up the profile. These are then punched out, using the output routine, on to the paper tape for feeding to the curve generator.

The type of curve generator varies with application. The medium general-purpose computer is most

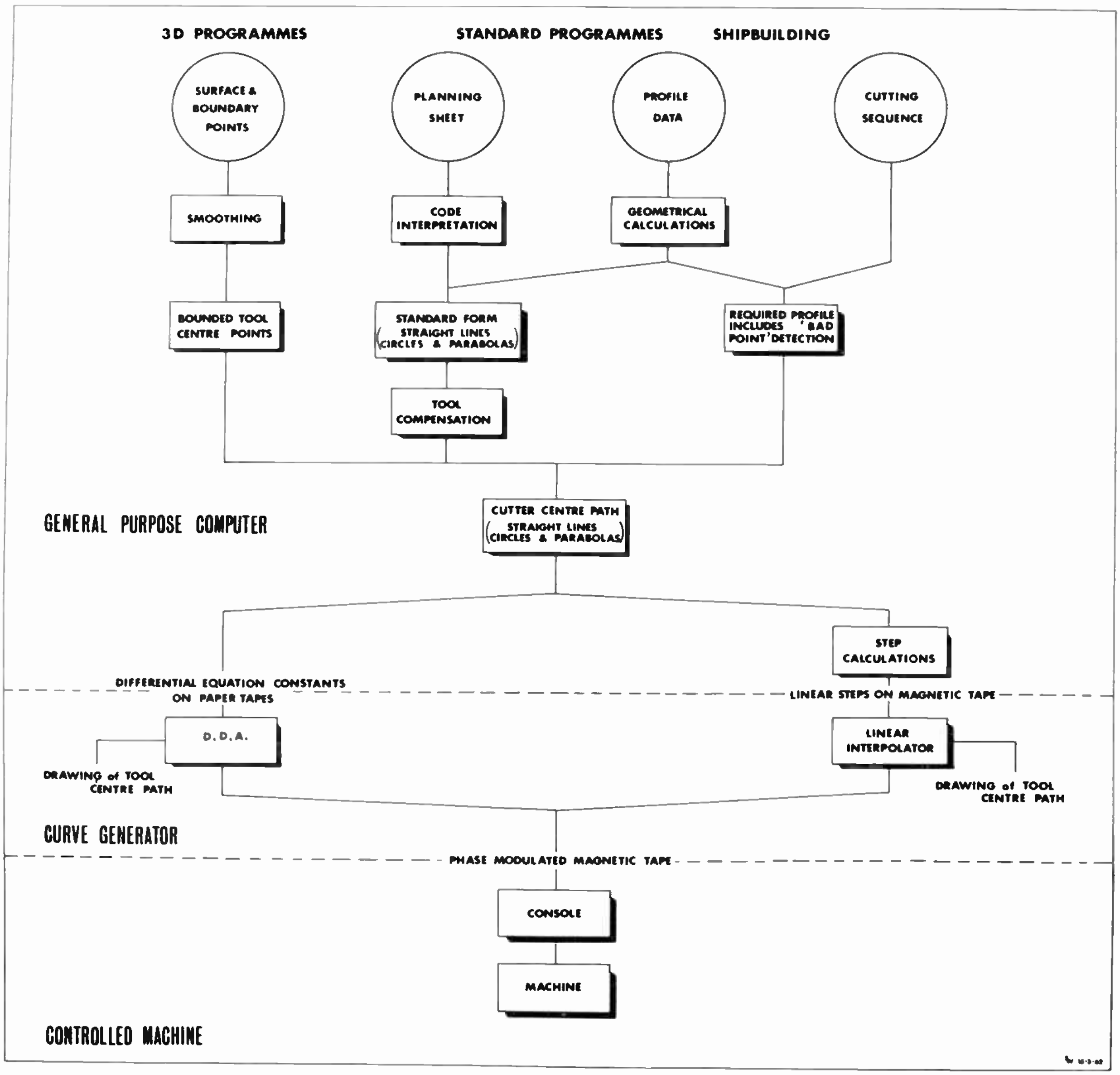


Fig. 2. Data processing subroutines.

efficiently utilized if it is coupled to a second order curve generator, i.e. one which calculates the continuous information from the differential equations of second degree curves. However, in cases where the processing system has only to feed a small number of machine tools and where the computing time is not at a premium, a more simple linear curve generator can be used. Instead of the cutter centre path information being specified as a series of straight lines, circles and parabolas, and punched out on paper tape, it is fed into another subroutine which reduces the profiles into a series of small straight lines. The length of the lines is chosen so that the curves are always to the required accuracy. These linear steps are recorded on computer magnetic tape and transferred to the curve generator which converts them into suitable control signals for recording on the machine tool control magnetic tape. The calculation of these linear steps in the general-purpose computer can be very time-consuming and this process is only justifiable when the amount of computation for machine tool control tapes is limited.

Although conventional engineering drawings are dimensioned with great accuracy, this is far from the case when the profile is an arbitrary curve; in such cases it is normal practice for the draughtsman to draw a smooth curve through the points and specify it by taking co-ordinate points from the curve. This information is of sufficient accuracy for a skilled craftsman to manufacture these curves or surfaces by conventional techniques, because he smooths out any unevenness but it is not good enough for numerical control where the dimensions must be exact. It is necessary to introduce a smoothing subroutine into the data processing program to take out the bumps and ripples in the information. This, of course, is only necessary in the cases of template cutting or surface milling where the information is obtained from the drawing board. If the surfaces are calculated in the computer then the smoothing is unnecessary. Likewise, in the flame cutting application to ships, profiles are specified from a scrieve board and it is necessary to remove bad points from the specification.

The ship-building planning procedure is described briefly in Appendix 1. It shows how the planning has been designed to be very simple and to be integrated into present manufacturing techniques.

5. Curve Generation

5.1. *Digital Integrators*

The path followed by the cutter centre is generated by solving, in a step-by-step fashion, the differential equations representing the required curve. This is accomplished by setting up and interconnecting digital integrators to simulate these differential equations in a similar way to analogue computers. The independent variable controlling these[2, 3, 4] integrators is arranged to increase in small incremental dt steps and generates incremental δx, δy and δz pulses on the output lines of the respective integrators.†

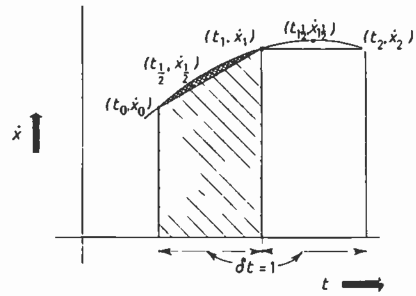


Fig. 3. Approximations for digital integration.

Various types of approximation can be used for this digital integration. An approximation which is very similar to the trapezoidal approximation used in graphical and numerical integration is commonly and very widely used and for most applications its accuracy is good enough. A typical curve showing the variations of velocity $\dot{x}$ against time t is shown in Fig. 3 in which it is required to integrate $\dot{x}$ with respect to t. In the case of trapezoidal approximation the integral

$$x = \int \dot{x}\,.\,\mathrm{d}t \qquad \ldots\ldots(1)$$

is approximated by

$$x = \frac{\dot{x}_0 + \dot{x}_1}{2} + \frac{\dot{x}_1 + \dot{x}_2}{2} + \ldots + \frac{\dot{x}_{n-1} + \dot{x}_n}{2} \qquad \ldots\ldots(2)$$

As the curves to be integrated are only known up to the last integration point, say $\dot{x}_0$, the mid-point values are predicted from the value of $\dot{x}_0$ and $\ddot{x}_0$. The incremental steps δt are made equal to unity to give the digital approximation as

$$x = \dot{x}_{1/2}\,.\,1 + \dot{x}_{1\frac{1}{2}}\,.\,1 + \ldots + \dot{x}_{n-\frac{1}{2}}\,.\,1 \qquad \ldots\ldots(3)$$

These approximations give errors and in the case of trapezoidal integration the error is represented by the cross-hatched area. As the digital integration takes the mid-point value of the $\dot{x}$ curve and not the mean value of $\dot{x}$ points on the curve, the error in this case is not identical but is of the same order as the trapezoidal case. By making the δt steps small enough, this error can be controlled to give the required accuracy at the expense of speed. If δt is chosen so that the incremental areas are less than 1 unit of x, the accumulation of these errors over complete curves can be kept less than 1 unit of x in most cases.

A block diagram of a digital integrator is shown in Fig. 4. It consists basically of two accumulators interconnected by a comparator. The R accumulator,

† A list of symbols is given in Appendix 2.

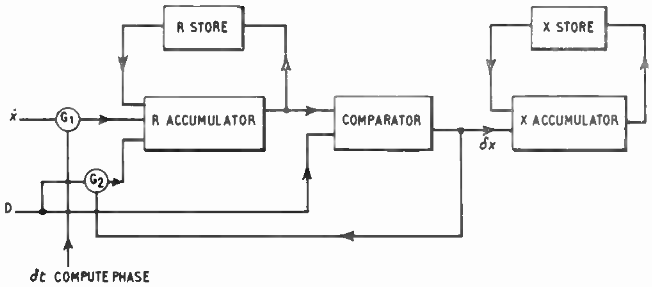


Fig. 4. Digital integrator.

$\dot{x} < D;\ \delta t_n = 1;\ R'_{n+1} = R_n + x_n$

(i) $|R'_{n+1}| < D$ $\quad \begin{cases} R_{n+1} = R'_{n+1} \\ x_{n+1} = x_n \end{cases}$

(ii) $|R'_{n+1}| > D$ and R_{n+1} + ve $\quad \begin{cases} R_{n+1} = R'_{n+1} - D \\ x_{n+1} = x_{n+1} \end{cases}$

(iii) $|R'_{n+1}| >$ and R'_{n+1} −ve $\quad \begin{cases} R_{n+1} = R'_{n+1} + D \\ x_{n+1} = x_n - 1 \end{cases}$

consisting of an adder and store, accumulates the fractional or remainder part of the summation process of equation (3), while the whole-number part is stored in the X accumulator, also consisting of an adder and store. Each δt compute pulse opens gate G_1, and allows the velocity $\dot{x}$ to be added to the R store. The new accumulated number in the store is compared with the scaling constant D and depending on whether it is greater or less than D, the R and X stores are modified appropriately. These conditions are shown in the Table associated with Fig. 4. In the case when D is greater than the contents of the R store, the comparator supplies a digital pulse to the X accumulator which increases it by 1 unit and opens the gate G_2 which subtracts D from the contents of R. The value contained in this store will now be less than D.

5.2. *Parabolas*

The parametric equations in t for a parabola in the x, y plane are

$$\left.\begin{aligned} x &= \frac{k}{2}t^2 + lt + m \\ y &= \frac{p}{2}t^2 + qt + r \end{aligned}\right\} \qquad \ldots\ldots(4)$$

This curve is generated from the differential equations

$$\left.\begin{aligned} \ddot{x} &= k \\ \ddot{y} &= p \end{aligned}\right\} \qquad \ldots\ldots(5)$$

with the correct initial conditions set in for the constants k, p, l, q, m and r.

As these parabolic curves are principally used for interpreting a smooth profile between a series of specified points, it is essential that the initial conditions of the fitted parabolas are chosen so as to generate a smooth curve. This is accomplished by drawing parabolas through the points so that adjacent intersecting parabolas have matched gradients. Figure 5(*a*) shows a parabola drawn through A and B. If the gradients at A and B are produced to intersect at 0, and the initial and final gradients taken as $\dot{X}_p$, $\dot{Y}_p$ at $t = 0$ and $\dot{X}_{p+1}$, $\dot{Y}_{p+1}$ at $t = 1$ respectively, then the initial conditions as shown in Fig. 5(*b*) are

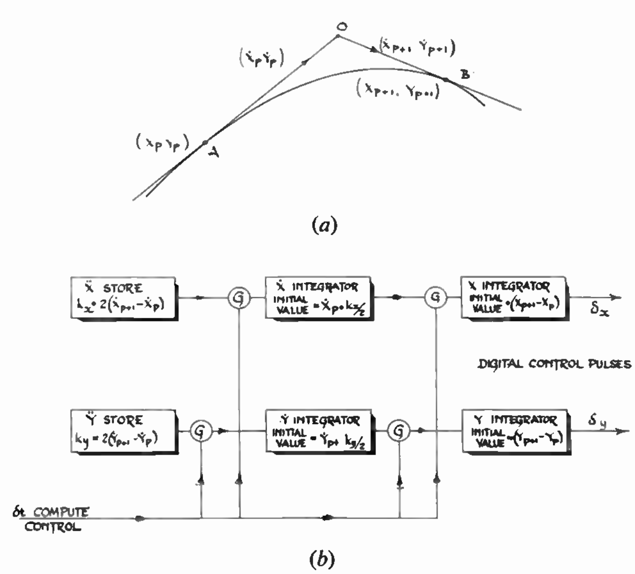
$$\left.\begin{aligned} k &= 2(\dot{X}_{(p+1)} - \dot{X}_p) \\ p &= 2(\dot{Y}_{(p+1)} - \dot{Y}_p) \end{aligned}\right\} \qquad \ldots\ldots(6a)$$

$$\left.\begin{aligned} l &= \dot{X}_p - k/2 \\ q &= \dot{Y}_p - p/2 \end{aligned}\right\} \qquad \ldots\ldots(6b)$$

$$\left.\begin{aligned} m &= X_p \\ r &= Y_p \end{aligned}\right\} \qquad \ldots\ldots(6c)$$

These equations are simulated by connecting two integrators in series for each X and Y variable (Fig. 5(*b*)), and the difference equations are

$$\left.\begin{aligned} \dot{x}_{n+\frac{1}{2}} &= \dot{x}_{n-\frac{1}{2}} + k.\frac{1}{2D} \\ \dot{y}_{n+\frac{1}{2}} &= \dot{y}_{n-\frac{1}{2}} + p.\frac{1}{2D} \end{aligned}\right\} \qquad \ldots\ldots(7a)$$



(*a*)

(*b*)

Fig. 5. Simulation of parabolic curves.

$$\left.\begin{aligned} x_n &= x_{n-1} + \dot{x}_{n-\frac{1}{2}} \cdot \frac{1}{D} \\ y_n &= y_{n-1} + \dot{y}_{n-\frac{1}{2}} \cdot \frac{1}{D} \end{aligned}\right\} \qquad \ldots\ldots(7b)$$

Each δt compute pulse opens the interconnecting gates and feeds the X and Y integrators with the incremental velocity components to the nearest digit while the acceleration components k_x and k_y are fed to the $\dot{x}$ and $\dot{y}$ velocity integrators; thus each δt pulse activates one step of each of the difference equations, generating δx and δy digital pulse trains from the output of the X and Y integrators. These pulses represent elemental control movements of the machine and when summed in the control servo trace out a parabola.

As the acceleration is constant for the parabola the integration approximation for velocity is exact. Similarly, because the velocity is rising linearly, the approximation for displacement will also be exact. The X and Y integrators will thus generate exactly the digital pulse rates to give a true parabola and the only errors will be due to rounding off. These errors are found in practice to be less than 1 digit.

5.3. *Straight Lines and Circles*

Straight lines are a special case of parabolic curves in which the acceleration components k and p are zero and are generated by the same method if only the X and Y velocity integrators are set with the initial velocities.

The differential equation for a circle with its centre at the origin is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{y} \qquad \ldots\ldots(8)$$

The difference equation for this circle is

$$\left.\begin{aligned} x_n &= x_{n-1} + y_{n-\frac{1}{2}} \cdot \frac{s}{D} \\ y_{n+\frac{1}{2}} &= y_{n-\frac{1}{2}} - x_n \cdot \frac{s}{D} \end{aligned}\right\} \qquad \ldots\ldots(9)$$

$$y_n = y_{n-1} - (x_n + x_{n-1}) \cdot \frac{s}{2D} \qquad \ldots\ldots(10)$$

Although the digital approximation is not exact it is found in practice that if the integrations are carried out in their serial fashion, errors in computation round a complete circle are less than 1 digit. The sign s in eqns. (9) and (10) determines the direction in which the circle is generated. The basic simulation of eqn. (9) is carried out by connecting the X and Y digital integrators in series with the output of the Y integrator connected back into the input of the X integrator. Equation (10) is simulated using a separate integrator to generate the true value of Y.

5.4. *Acceleration and Velocity Control*

As well as generating the correct form of profile it is necessary to control the velocity at which these curves are generated to give the correct feed rate. Also the machine tool servos cannot accelerate immediately to the cutting velocity and an acceleration restriction must be built into the process. If the machine table starts from rest the table velocity should rise linearly to the required feed velocity and remain constant until it nears the finishing point when it should fall to zero with constant acceleration (Fig. 6(*a*)). In cases where the curve is continuous the velocity would not fall at this finishing or change point but would continue at the feed velocity. The feed rate need only be accurate to 5% of its value and analogue techniques are employed for reasons of simplicity.

The velocity control is obtained by controlling the rate at which the curve is computed (Fig. 6(*b*)). The compute C_p or δt pulse rate, instead of being constant, is generated from a special pulse rate register R. The rate at which it generates overflows depends on the product of the pulse rates P_r and the velocity constant V_k which is fixed by the feed rate. The pulse rate P_r is generated from analogue feedback signals to give the correct r.m.s. velocity and to limit the acceleration. When the machine has reached the feed velocity the acceleration restriction has no control and the pulse rate P_r is controlled from the component output velocities. This takes their analogue values and forms the voltage

$$\frac{D}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}}$$

This analogue voltage is fed to the voltage/pulse rate converter which generates a pulse rate proportional to it. The output pulses operate gate G feeding V_k into the R register and giving an output pulse overflow rate

$$C_p = V_k \cdot \frac{D}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} \qquad \ldots\ldots(11)$$

and the r.m.s. value for the components $\dot{x}$, $\dot{y}$ and $\dot{z}$ feed velocities

$$v = C_p \cdot \frac{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}}{D} = V_k \qquad \ldots\ldots(12)$$

i.e. the required feed rate.

The r.m.s. velocity feedback control is done in this way so that the analogue control signals are always of the same order and never very small, enabling accuracy of control when the feed rate is very small.

The machine acceleration which is represented by the acceleration number is decoded into a current by the acceleration decoder. This current is fed into a capacitor, the voltage across which controls the pulse

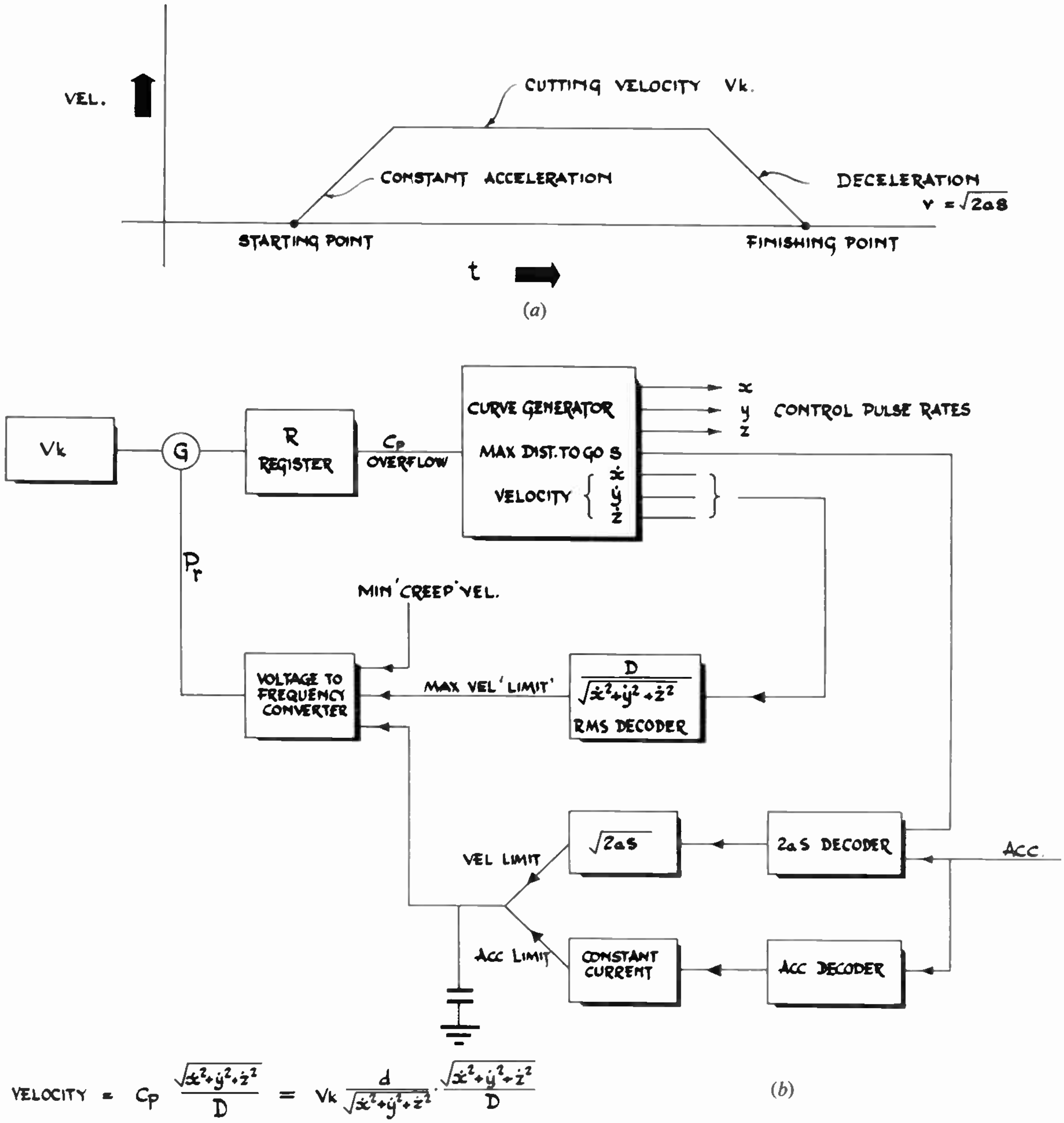


Fig. 6. Acceleration and velocity control.

rate of the voltage-to-pulse rate converter until the r.m.s. cutting velocity is reached. The pulse rate P_r will thus increase linearly and so will the output of the R register until the feed-rate velocity is reached, after which the r.m.s. velocity control will take over.

The run-down velocity control need not be very exact and the distance to the finishing point S is taken as the maximum of the X, Y and Z component distances to this point and decoded in the $2aS$ decoder. The acceleration which need only be controlled within $\pm$ 50% is represented by a single binary digit. The value of this binary digit weights S to give an output voltage of $2aS$. This is then fed to an analogue square root device to generate $\sqrt{(2aS)}$. The output voltage from this generator is used to restrict the input control to voltage/pulse rate converter when $\sqrt{(2aS)}$ falls below the feedback voltage from the r.m.s. decoder. Thus, the value from $\sqrt{(2aS)}$ generator will be high to start off with but as the finishing point is reached it will fall to a low value when it will take over from the r.m.s. velocity control. At this point the pulse rate will gradually fall to zero giving the required control on the output control pulse rates as the finishing point is reached.

6. Tape Link

6.1. *Digital to Phase Converter*

The information generated from the three control channel integrators is in the form of incremental digital pulses. These control pulse trains must be converted to phase-modulated information for recording on the magnetic tapes. The reference channel square waves are generated by counting down by 100 a square wave generated from a 260 kc/s oscillator (Fig. 7) giving a reference square wave of 2·6 kc/s. The control channel square wave is generated by a similar procedure. In this case the divide by 100 counter is fed from the incremental control pulses from the d.d.a. as well as from the high frequency oscillator.

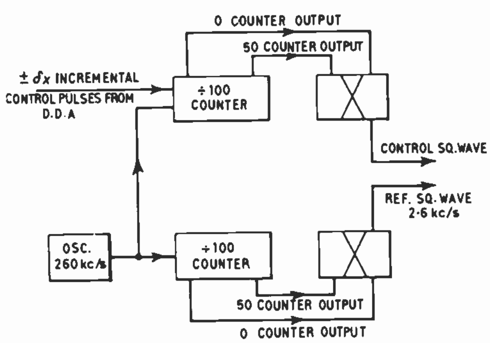


Fig. 7. Digital to phase converter.

The two separate trains of pulses are interlaced so that the counter does not receive two pulses simultaneously. If no incremental pulses are being received the control square wave will be 2·6 kc/s and will remain at a constant phase difference with respect to the reference square wave. If an incremental δx pulse is added into the counter the phase of the control square wave is advanced by 1/100 of a cycle with respect to the reference square wave. Likewise, if a negative incremental pulse is received the phase is retarded by 1/100 of a cycle. In this way the control square wave is phase modulated with respect to the reference square wave and follows exactly the changes in the incremental pulse rate coming from the d.d.a.

6.2. *Error Correction*

Magnetic tapes may suffer from imperfections causing dropouts in the recorded signals. These dropouts are corrected for by an error detecting technique. The signals are read after writing and all tapes with two consecutive dropouts are rejected. It is found that the majority of tapes will pass this test. At the machine a special flywheel-synchronizer injects pulses where the single dropouts occur. This circuitry is designed so that it can cope with at least two consecutive dropouts on the tape. Occasional noise pulses may also interfere with the signals read from the tape. Special masking pulses are generated from the control signals; these pulses inhibit the reading of any further control signal until just before its predicted time of arrival and prevents any noise pulses generated in these gated intervals from being detected as false control signals. This gives a link with very wide operating margins and free from dropout and noise troubles.

6.3. *Machine Functions*

As well as recording the control signals to move the table of the machine to generate or to trace out the profile, it is also necessary to record machine function signals such as coolant on/off, spindle on/off. This information is recorded as a four-bit number, one bit per channel giving up to 15 special machine function signals. Each bit consists of an 8-cycle phase disturbance recorded on the tape in between the control phase cross-overs. Special filtering circuits exist in the control console for separating the control information from the machine function signals after it is read off the tape.

It is most important that the machine function controls should give faultless operation. The reading circuit for these signals must detect at least four of these signals before it operates. Noise pulses will therefore not affect the operation and neither will dropouts as at least eight signals are recorded on the magnetic tape. The signals are also checked as they are processed through the system. The subroutine which is used to transform these signals on the planner's program tape to the signals for punching out and feeding to the curve generator has a special check routine. The program of the general-purpose computer weights the bits of these machine function signals and accumulates them in its summing register. A summation check is punched out with each machine function code and is used to check that the correct signals have been recorded on the magnetic tape.

7. Curve Generator

7.1. *Logical Description*

A block diagram of the curve generator which has been designed to work with a *Pegasus* general-purpose computer is shown in Fig. 8. In order to give clarity, only one control channel is shown along with the reference channel. The curve generator, which is used to record tapes for 3-dimensional milling, has 3 control channels; one for each axis of the machine and the reference channel.

The paper tape from the *Pegasus* computer is fed into the curve generator section by section. After the paper tape information is read, decoded and assembled in words, it is stored in the buffer store ready for transfer to the digital curve generator when required.

When the tape is first read, the magnetic tape deck is started up and records continuously during the whole of the computing process. The digital curve generator computes continuously the path of the cutter centre, with its rate being controlled by the velocity and acceleration control unit to obtain the correct feed rate. The feed rate and acceleration constant of the machine are fed in initially on the paper tape as the first block of information. The *X*, *Y* and *Z* digital pulse rates from the curve generator are fed into the digital-to-phase converter which generates the correct control phase information for recording on the magnetic tape.

During the time the digital curve generator is computing a section of the profile, the section immediately following is being read from the paper tape and stored in the buffer stores. In order to cater for very short sections when the digital curve generator has completed the profile computation, before the paper tape reader has had time to read in the next section, a double set of buffer stores is used.

The signals are read off the magnetic tape as they are recorded and, after amplification and wave form shaping, are fed to the checking equipment. The presence of phase cross-over pulses on the tape are detected using two amplitude threshold levels, a high and a low one. The difference in these two sets of phase cross-over pulses for each channel is accumulated in the stores of the tape quality check and displayed to the operator. The high level threshold phase cross-over which is the most critical case for dropouts, is also fed to a double dropout detector which detects if more than one consecutive dropout is obtained at one time. If a large number of dropouts or write-ins are obtained, e.g. 10 or more, or if more than one consecutive tape fault occurs, a fault is indicated and the tape is rejected. The write/read check sums the difference between the total number of cross-overs recorded on the tape and the total read off the tape. At the end of each recording the operator checks with the visual display to make sure this difference is zero. The machine function control (m.f.c.) signals are also decoded off the magnetic tape and summed and compared with the m.f.c. sum check from the paper tape.

The paper tape code, each numerical character of which is divisible by three, is also checked for divisibility when the numbers are read into the buffer stores. Immediately a check failure occurs, the type of fault is indicated to the operator and the curve generator is shut down.

It is customary with machine tool programs to start always from a datum and return to it when the cutting is complete. This gives an additional overall check on the operation of the control equipment and sets the machine table in the correct position for the next operation. The datum check is used in a similar way to make sure that no errors have occurred in the *Pegasus* program or in the computing of the profiles by the curve generator. As the datum count-check accumulates the difference of the control and reference channel cross-over pulses and not the incremental digital pulses, the operator must first of all check that the accumulated numbers in the x, y and z datum stores are zero at the end of the program. This in itself is only capable of determining phase accuracy to the nearest half cycle. He will then be able, by means of a phase comparator to check, at the same time, that the *X*, *Y* and *Z* signals are in zero phase with respect to the reference, so giving a check on datum return to the nearest bit size.

7.2. *Electronic Circuits*

The electronic circuits are housed in a U-shaped desk high console (Fig. 9). The plotting table,

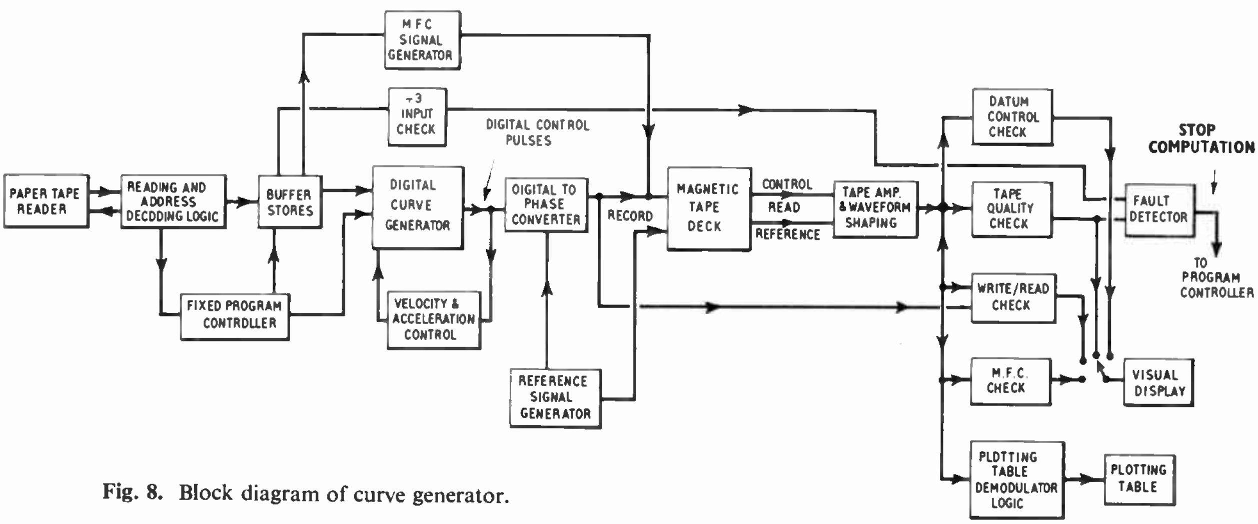


Fig. 8. Block diagram of curve generator.

Fig. 9. The curve generator.

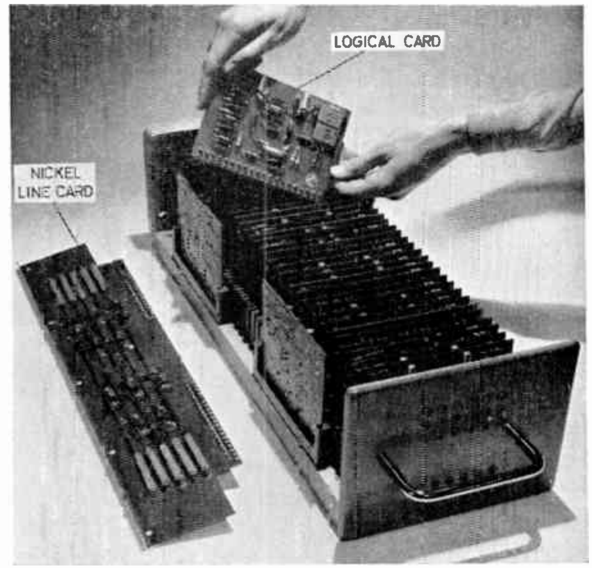


Fig. 10. Plug-in tray unit of the curve generator.

tape deck and paper tape reader are situated on top of the electronics and are arranged to be within easy reach of the operator. The electronic components are assembled on printed circuit cards 5 in × 4 in. These cards are housed in a 2 ft long plug-in tray unit having a capacity for 36 plug-in cards (Fig. 10). The nickel delay lines which are used for word storage are housed in long cards which plug into sockets along the side of the tray.

Computing is carried out serially with a digit rate of 500 kc/s and a word length of 30 digits. The equipment employs transistors throughout and computation is carried out using circuitry very similar to the SEAC computer in America.[5] In this method the logical operations are carried out using conventional diode gates and the pulses dynamically re-shaped using a transistor amplifier. This amplifier is a transformer-coupled blocking oscillator which is triggered by the input signal and in which the feedback maintains the output to generate a re-shaped pulse. The timing of this pulse is controlled by the timing system which is used to sample the input and switch the feedback off.

A four-phase synchronous timing system is used and each digital pulse is re-shaped after every logical operation and delayed by a quarter digit period. Electromagnetic delay lines are used for single digit storage, and for obtaining the short delays used in the correct phasing of the digital pulses, when required, before feeding into the logical gates. The electromagnetic delay is a lumped LC transmission line in which the capacitance is obtained from a bifilar-wound coil. The capacitance between the two wires of this coil give the lumped C for the line by connecting one wire of the winding to earth, and the other wire to the adjacent inductance coils. The lines are wound on a special machine which measures the resonant frequency as the wire is wound on to give an accurate control of the phase delay.

The logical circuitry was designed to allow for $\pm$ 20% tolerance in the value of resistive components and $\pm$ 10% variation in supplies, and to operate up to temperatures of 55° C with the critical parameters of the transistors set at their extreme values. The rules used for connecting the logical units together were kept to a minimum and as simple as possible. The loading rules were purposely derated from the design values to improve the working margins and reliability of the equipment. In the case of the analogue rate control, the maximum variation with temperature in the transistor parameters could not be catered for in the design, and the transistors were housed in temperature-controlled heat sinks. Silicon transistors were used in all cases where leakage current was critical. A marginal checking feature was built into the positive and negative signal bias lines to enable preventive maintenance to be carried out, and to weed out marginal effects.

7.3. *Input Output Equipment*

7.3.1. Tape reader

A high-speed paper tape reader is used for reading the *Pegasus* output tape. An optical reading head with solar cell light detectors reads the five-hole paper tape. A simple fast-operating electromagnetic brake, which clamps the tape between two metal plates, enables a very fast stopping time to be obtained. When the brake clamp is applied initially, the paper tape slips round the drive capstan until it is disengaged to prevent paper wear. The maximum paper-tape reading speed is 1000 characters per second for the reader to stop within one character. To improve the reliability it is run at half speed which is fast enough for this application.

7.3.2. Magnetic tape deck

The magnetic tape deck records ¼ in wide tape at 100 in/s and 50 in/s giving record/replay ratios of 26 : 1 and 13 : 1 respectively. The lower record speed is used in cases where the best surface finish is required; by sacrificing computing speed to digit size, increased accuracy and linearity are obtained.

The magnetic tape deck controls which operate during the recording period are completely transistorized in order to cut down to a minimum the electrical noise generated in the vicinity of the computer.

7.3.3. Plotting table

The plotting table has been designed to operate at a high speed to allow a drawing of the part to be traced as the magnetic tape is recorded. The machine reproduces plan (30″ × 10″) and elevation (30″ × 7″) views of the cutter centre path. The X channel motor drives the carriage on which the Y and Z travelling pens are attached. Each axis is driven by an M-type electrical motor which is fed with four-phase control signals. A demodulator removes the carrier component from each control signal and generates a suitable signal for driving the M motor. The machine is designed on kinematic principles to reduce drive friction with the carriage weight kept to a minimum to give the best dynamic performance. The M-motor drive is transmitted via reduction gear boxes to the carriage assembly by means of non-slip steel tapes. The maximum plotting speed is 2 in/s and overall accuracy is 0·005 in.

8. Data Processing Costs

Cost comparisons with conventional techniques are very difficult to make because of the wide variety of factors which have to be taken into account, depending on each particular application. However, numerical control costs can be readily estimated. They are to a first approximation dependent only on the number of individual profile sections making up the complete cutting operation. Some cost figures for small milling machines are shown in Table 1. Experience on controlled machines has shown that with this type of machine the average profile section length between change points is about 2 inches. This fact allows the various items in the process to be readily estimated as shown in Table 2.

Table 1

Cost figures for small milling machines

ITEM	Present Cost		Future Cost
	per section (of profile)	per foot (of profile)	per foot (of profile)
CONVENTIONAL.			
Simple milling operations.		5/-	
Complicated milling operations.		15/-	
Templates.		£5	
Surface milling (Turbine blades).		£1	
COMPUTER CONTROL.			
Planning.	9d	4/6	2/3
Computer.	8d	4/-	2/-
Machining, Aluminium.	2⅔d	1/4	8d
Mild steel.	8d	4/-	2/-
Total cost, Aluminium. 1 off	1/8	9/10	4/11
10 off	5d	2/2	1/1
Mild steel. 1 off	2/1	12/6	6/3
10 off	10d	4/10	2/5

Table 2

Average profile section length	2 inches
Planning: 3 minutes per section at a cost of 15s. per hour including overheads	9d. per section
Computer time: 4 seconds per section at £30 per hour	8d. per section
Machine running costs: £4 per hour including overheads and assuming 75% utilization.	
(a) Aluminium machining at 12 in/min and average time between change point of 10 seconds	2⅔d. per section
(b) Mild steel machining at 4 in/min and average time between change point of 30 seconds	8d. per section

The planning and computer costs per part produced depend on the batch size and when the batch size is greater than one then these costs must be divided among the batch.

The figure for conventional milling was taken for a batch of 1–10 with a 20% utilization of the machine running at a cost of £1 per hour. The feed rate was taken at 4 in/min. The figures for surface milling and template manufacture were taken from particular examples. These conventional figures can vary widely and should only be taken as a rough guide for comparative purposes.

From these figures it is seen that the planning and computer costs for one-off jobs are the major part of numerical control production costs and are comparable with conventional production. However, for small batches of 5–20, these high costs are proportionately divided among the batch, reducing considerably the computer costs of numerical control. It is in this sphere where numerical control is most profitable and competitive with other forms of production, and one-off-type components tend to be restricted to the more complicated and expensive milled components, templates and dies. By improving the planning and by use of faster computers, the position should change in favour of numerical controlled machines. The last column in Table 1 indicates how these costs should fall in the near future when full use is made of second generation computers. It is not too unrealistic to assume that costs for any computing operation in these new computers will fall by at least 4 times. If the planning is improved to make it simpler and to reduce the planning costs it will probably require a computer subroutine, twice as complex as before. The net effect will be a reduction in the computer costs by a half and the improved programming for the planner will likewise reduce the planning costs by one half. The continuing development and improvement in machining techniques, better machine geometry and improvement in the servo system will also enable the machine feed rates to be increased by a factor of 2, reducing the machining costs by a corresponding factor. These future developments will make numerical control more and more profitable and competitive with conventional techniques.

A factor which will probably play a very important part in this economic comparison will be the ever-increasing shortage of skilled labour. This will be reflected in increased costs in conventional techniques of manufacture and will force manufacturers more and more into use of numerical techniques.

9. Future Trends

Future developments in the numerical control field will be mainly concentrated in improving the data processing and widening its application. Considerable work is going on in the use of computer programs with applications to die sinking. Although some progress has already been made and surface milling programs already exist from which a limited variety of dies can be manufactured, there are still many problems to be solved in this field.

The Americans have devoted considerable effort to these machine tool programs, such as Auto-Prompt and APT; these tend to be too comprehensive and have become somewhat unwieldy in use. They would seem to be extremely expensive to use and sufficient thought has not been given to the high intelligence required of planners, operators and users, e.g. surfaces are specified by mathematical shapes which draughtsmen and planning engineers will find difficulty in understanding.

This development will go hand and hand with the use of computers in design. The manufacture of turbine blades is an application where this association has already started. Here the blade is basically designed in the computer and drawings produced of experimental blades which will be used in conjunction with computer programs to check the design. Control tapes are recorded for the manufacture of the experimental blades. When tests on these experimental blades show up some weaknesses the design is modified and corrected and a new set of blades manufactured for approval. After this process has been completed the manufacture and production of blades will go straight ahead provided sufficient controlled machines are available. This will reduce enormously the time between the design and production stages of components.

10. References

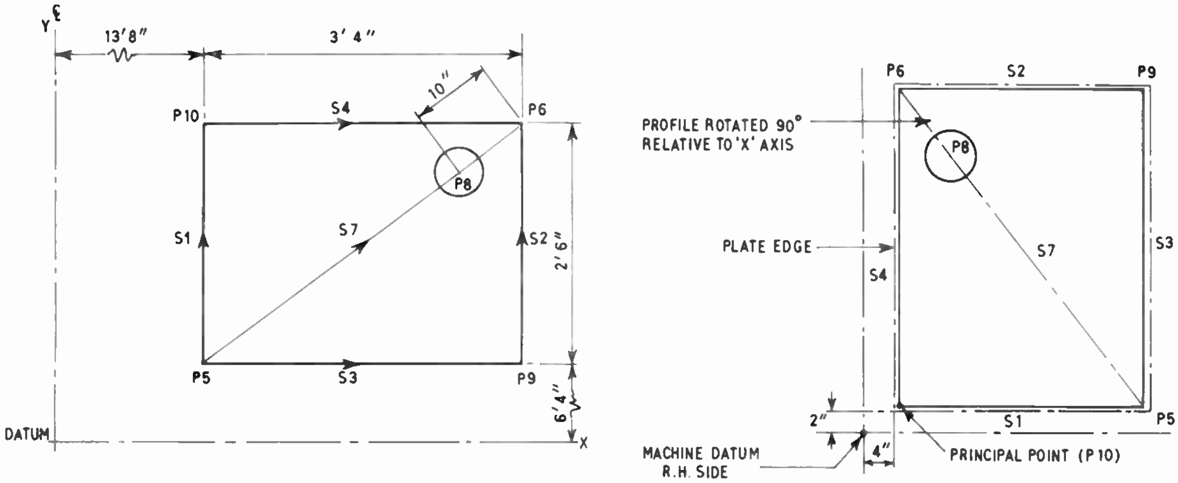
1. D. F. Walker, "Ferranti Mk. IV continuous path phase control system", *Int. J. Machine Tool Design and Research*, **3**, No. 1, pp. 61–73, January–March 1963.
2. R. E. Sprague, "Fundamental concepts of the d.d.a. method of computation", *Mathematical Tables and Other Aids to Computation*, **6**, No. 37, p. 41, January 1952. (Preliminary description of d.d.a. principles and comparisons with analogue d.a.'s.)
3. G. C. Toothill, "The incremental digital computer (digital differential analysers)", *Process Control and Automation*, **5**, Nos. 10 and 11, September and October 1958.
4. K. Millington, "An experimental incremental computer", *The Radio and Electronic Engineer* (*J. Brit.I.R.E.*), **25**, No. 5, pp. 461–73, May 1963.
5. S. Greenwald, R. C. Haveter and S. N. Alexander, "SEAC", *Proc. Inst. Radio Engrs*, **41**, pp. 1300–13, October 1953.

11. Appendix 1

In order to give clarity to the explanation of the planning system used in the flame cutting application of ship's plates, a simple oblong plate with a hole in the diagonal (Fig. 11(*a*)) has been taken to illustrate the method of planning. The planner specifies the geometry by filling up a profile data sheet (Fig. 11(*a*)) and then a cutting sequence chart (Fig. 11(*b*)) which gives all the information necessary for cutting the plate.

All points, lines and curves on the steelwork drawing are addressed with numbers. In Fig. 11(*a*), the 4 points of the plate are P10, P5, P9 and P6 and the lines adjoining these points S1, S3, S2 and S4. The points and lines are then defined by the co-ordinate distances from the reference axis or from previous defined information. For instance, the first line of the profile data sheet reads "S1 is a vertical line 13′ 8″ from the Y axis". S3 is defined in a similar way but with H in place of V in Column 1 of definitions as it is a horizontal line. Lines parallel to one another are specified as in line 2 of the profile table which reads "S2 is a parallel line to S1 lying to the right of it at a distance of 3′ 4″ when looking along the direction of S1". S4 is specified similarly, but as, in this case, S3 is to the left of S4, L is entered in place of R in the last column of the table.

Point P5 is the intersection of S1 and S3 and is defined by writing S1 and S3 in columns 2 and 3 of the definitions table. The diagonal line S7 is specified in a



REQ	DEFINITIONS 1	DEFINITIONS 2	DEFINITIONS 3	CO-ORDINATES X FT	X INS	X FRAC	Y FT	Y INS	Y FRAC	RADIUS OR DISTANCE FT	INS	FRAC	FROM
LOCATION L19													
S1	V			13	8	0							
S2										3	4	0	RS1
S3	H				*		6	4	0				
S4										2	6	0	LS3
P5		S1	S3										
P6		S2	S4										
S7		P5	P6										
P8	D	S7								0	10	0	P6
P9		S2	S3										
P10		S1	S4										
END													

(*a*)

LOCATION	L	19		
PRINCIPAL POINT	P	10		
DATUM CO-ORDS OF PRINCIPAL POINT	X	0	4	½
	Y	0	2	½
ANGLE TO 'X' AXIS	B	+90		
KERF WIDTH INS.	K	0·080		
CUTTING SPEED INS/MIN	R	20		

POINT	CURVE (+OR-)	SUB-ROUTINE			
P8		C1	B	6	
P6		PIERCE			
P10	-S4				
P5	-S1				
P9	+S3				
P6	+S2	RAISE			
DATUM					
END					

(*b*)

Fig. 11. Planning of the cutting sequence for flame cutter.

similar manner by writing P5 and P6, the point through which it passes, in columns 2 and 3 of the definitions. The centre of the manhole P8 is specified on the 8th line which reads "P8 is a point on the diagonal line S7 at a distance -10 from the point P6".

The sequence in which the plate is to be cut is detailed in the cutting sequence chart (Fig. 11(*b*)). At the top of the chart the following general cutting details are filled in:

The cutting speed (20 in/min)

The principal point (P10)

The co-ordinate of the mechanical datum for setting up the job referred to the principal point P10.

The kerf width or flame diameter.

The plate profile often has to be oriented to the X axis by an angle B and possibly fitted in with other plate profiles in order that it can be cut from standard plate sizes with the minimum of scrap. In this particular case the plate has been oriented through 90 deg.

The planner then describes each cutting operation in sequence, filling up the planning sheet line by line. The second line reads "move to point P6 from the previous point", which in this case was P8. Similarly, the 3rd line reads "move to P10 from previous points along line S4 in the negative direction. In the flame cutting application the flame must be ignited with the nozzle in the right position to start the cutting of the profile. This is done by writing "pierce" instructions on the planning sheet which stops the magnetic tape deck on the control console of the machine, lowers the nozzle, switches on the gases and ignites the flame. When the flame has been successfully ignited, the tape deck is switched on again and the cutting head traces out the complete profile. When the cutting of the profile has been completed the "raise" instructions are written on the planning sheet. This extinguishes the flame and raises the nozzle before it returns to datum.

As well as cutting out the profile, the machine has to cut a large number of notches where the plate butts against the framework of the ship and the various manholes to allow access between ship's compartments. The cutting sequence for these profiles is done automatically in the computer program and all the planning engineer has to do is to call up the correct subroutine, and specify its major axis and its centre.

The manhole P8 is defined in the first line of the cutting sequence chart which reads "cut manhole C1 which in this case is a circle of radius 6″ with its centre at P8". In flame cutting it is convenient in many cases to leave enclosed areas which have been cut out attached to the main plate rather than let them fall on the floor. This is done by leaving bridge pieces between the part being cut out and the main plate. This procedure is automatically carried out by writing a B in the second subroutine column. Where bridge pieces are not required P is written signifying a phase cutout.

12. Appendix 2: List of Symbols

$x\ y\ z$	co-ordinate values of the generated profile
$\dot{x}\ \dot{y}\ \dot{z}$	component velocity values of the generated profile
$\ddot{x}\ \ddot{y}\ \ddot{z}$	component acceleration values of the generated profile
$X\ Y\ Z$	initial co-ordinate values of the generated profile
$\dot{X}\ \dot{Y}\ \dot{Z}$	initial component velocity values of the generated profile
t	time variable
$\delta x\ \delta y\ \delta z\ \delta t$	incremental change in $x\ y\ z$ and t
k, l, m, p, q, r, s	equation constants
C_p	rate of computation of the curve generator
P_r	voltage to frequency output pulse rate
V_k	feed rate constant for machining
S	distance to the finishing point
a	maximum permissible acceleration for the machine tool
D	digital integrator scaling constant.

Manuscript first received by the Institution on 12th December 1962 and in final form on 18th May 1963. (Paper No. 872/C64).

POINTS FROM THE DISCUSSION

Mr. A. H. Erlund: Would the author please explain why the computer itself is not used to plot the curve, which is required as discrete points, instead of using a curve plotting machine.

The author (*in reply*): If the computer was used to produce a drawing directly as a series of points, a smooth curve would have to be drawn through these points, or, alternatively, they would have to be very close together for a reasonable reproduction of the cutting profile path. Both methods are not very satisfactory and the latter method would be very time-consuming in computer time.

It is thus necessary to use a curve plotting machine which will interpolate information supplied by the computer and draw smooth curves through the specified points.

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".

AUTOMATIC ASSEMBLY

Selecting each of various components of a sub-assembly, as it is manufactured, checking its individual specifications and setting up an appropriate test circuit lead to a programming problem. Two French engineers describe a relatively simple solution making use of magnetic matrices, each having a rectangular hysteresis loop. Different possible programs are expressed by different paths followed by a number of wires through different magnetic circuits. The equipment is designed for a total capacity of 40 groups of 36 combinations, each one made up of 79 binary numbers—a total of approximately 100 000 binary numbers.

"Programming and sub-assembly automatic control apparatus by means of magnetic matrices", M. Chaleat and B. Loscul. *L'Onde Électrique*, **43**, pp. 802–5, July/August 1963.

STEREOPHONIC LISTENING

In listening to stereophonic reproduction by means of two loudspeakers, there appear to be certain anomalies of location of the sound sources. Some of the reasons for this are the acoustic nature of the listening room, the quality of the loudspeakers and the fundaments of the recorded sound. Experiments have been carried out by a French broadcasting engineer using an anechoic chamber to eliminate, as far as possible, the causes of misconception. These experiments have shown that the mere coexistence of two 'real' sources within the hearing range of the listener is the cause of local disturbances in the frequency spectrum which affect the apparent position and the tonal quality of the sound sources.

"Hearing anomalies in stereophonic listening", M. L. Chateney. *L'Onde Électrique*, **43**, pp. 806–12, July/August 1963.

MICROWAVE MEASUREMENTS

It has been shown that perturbation methods of measuring dielectric constants at microwave frequencies are only convenient for relatively small values of dielectric constant and rod radius. Two Swedish engineers have recently shown that, by using a circularly cylindrical cavity with a dielectric sample as a central rod, the results are valid for high ϵ-values. Curves and tables, based upon an exact calculation of the resonant frequency behaviour of the TM_{010} mode as a function of ϵ, are presented for determining ϵ and tan δ from measurements of the resonant frequency and the Q-value of the cavity.

"Cavity method for measuring dielectric constants at microwave frequencies", P. Hedvall and J. Hägglund. *Ericsson Technics*, **19**, No. 1, pp. 89–96, 1963.

EDDY-CURRENT CALCULATION

The problem of finding an exact mathematical solution for a static two-dimensional scalar magnetic field can be overcome by using the relaxation method. Due to the complexity of the calculations it has been found necessary to use a digital computer. Time-changing quantities cannot be solved by ordinary relaxation but if the quantities vary sinusoidally it is possible to eliminate time by introducing complex quantities. Two methods worked out by Swedish engineers seem to give about the same optimum value in a way which is simpler than that used previously. Vector fields in two dimensions can also be calculated by relaxation on account of the fact that every vector field in two dimensions corresponds to a two-dimensional scalar field.

"Magnetic eddy-current fields calculated by complex successive overrelaxation", J. Ehrenborg and J. E. Sigdell. *Ericsson Technics*, **19**, No. 1, pp. 29–56, 1963.

RING FILTERS FOR MILLIMETRE WAVES

Ring filters constitute a suitable form of frequency branching networks in the millimetre-wave region. As resonators they contain rings of rectangular waveguide which are coupled with one another and with the input and output guides via directional couplers (rows of holes). In a recent German paper it is shown that the selectivity increases with an increasing number n of rings. With appropriate coupling the attention function becomes a Chebyshev polynomial of degree $2n$.

"Ring filters as frequency-branching networks for the millimetre-wave region", A. Jaumann. *Nachrichtentechnische Zeitschrift*, **16**, No. 6, p. 297, 1963.

CIRCULAR POLARIZATION

Studies have been made in France of the field received by a horizontally polarized aerial placed in a wavefront of any kind of polarization. It has been shown that where the polarization of the transmitted wave is circular the repetition rate of the received signal is equal to the frequency plus or minus the angular speed of aerial rotation. This is a criterion of circular polarization and may be applied to aerials which are independent of frequency. An aerial of this type should have circular polarization in every direction that it radiates including its axis, and a change of frequency should thus equal an angular rotation.

"Circular polarization and aerials independent of frequency", C. Ancona. *L'Onde Électrique*, **43**, No. 434, pp. 503–7, May 1963.

SUBJECT INDEX

Papers and major articles are denoted by printing the page numbers in bold type

INDEX OF PERSONS

Names of authors of papers published in this volume are indicated by bold numerals for the page reference.

Contributors to discussion are indicated by D

INDEX OF ABSTRACTS

This index classifies under subject headings the abstracts published throughout the volume in "Radio Engineering Overseas"

JOURNALS FROM WHICH ABSTRACTS HAVE BEEN TAKEN DURING THE SECOND HALF OF 1963

Acta Electronica (France)
Annales de Radioélectricité (France)
Archiv der Elektrischen Übertragung(Germany)
Canadian Electronics Engineering
Electronic Applications (Holland)
Elektrotekniski Ventnik (Yugoslavia)
L'Onde Électrique (France)
Mitsubishi Denki Laboratory Reports (Japan)
Nachrichtentechnische Zeitschrift (Germany)
N.E.C. Research and Development (Japan)
Philips Technical Review (Holland)
Proceedings of the Institution of Radio Engineers, Australia
Slaboproudý Obzor (Czechoslovakia)
Telecommunicatii (Hungary)
Tijdschrift van het Nederlands Radiogenootschap (Holland)