THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 32 OCTOBER 1966 NUMBER 4

INSTANT INFORMATION

INFORMATION is all the more valuable if it is both up to date and quickly available. It is not possible to provide such a service for scientists and engineers through the medium of a concordance or abstracts since the time involved in selection and writing may cause the resultant information to be out of date.

It is a sad reflection on what planners are pleased to call 'a fast moving technological age' that a system for providing up to date information in any field of learning was, in fact, propounded twenty years ago this month!*

The 1946 proposal lacked the strong support of a non-profit-making and research-minded organization. The formation of the National Electronics Research Council with membership drawn from Government, industry and the universities, provided such an organization and announced in 1964 that it would back research into the whole problem of the dissemination of information. Indeed, it lodged with the British Government in October 1964 a request for financial support of the investigation.

Progress reports on the first two phases of what is now known as the S.D.I. project have appeared in the *N.E.R.C. Review*. The Director for the Office of Technical and Scientific Information has confirmed that 'this is the first time in Britain that an S.D.I. service has been planned for the benefit of the scientific public as a whole and not confined to the members of the sponsoring organization.'

Mechanized information retrieval systems have, of course, attracted attention in several other fields of learning. The Chemical Abstracts of the U.S.A. provides a current awareness information service for chemists; in nuclear science the U.S. Atomic Energy Commission which has hitherto provided a restricted abstract service, propose to initiate a mechanized service in the near future. In May 1965 an information service for physicists based on computer manipulation of data obtained from published literature on physics was initiated by the Institution of Electrical Engineers, and the cost of this work is now wholly to be financed by a Government grant. More recently the Medical Research Council, again with a grant from the Government, is operating a small research project FAIR (Fast Access Information Retrieval).

These and other systems have been described in *N.E.R.C. Review*, but all of them have been tailored to the individual requirements of a company or specialist discipline. The S.D.I. system proposed by the National Electronics Research Council aims at providing immediate current information. Because it is carefully directed towards the actual requirements of users the system will build up a high retrieval potential. Although electronics alone is the subject used for the experiment, S.D.I. can be applied to every other form of learning.

As with almost every research project, finance is a great stumbling block; although a very generous contribution has been offered by the British Government an equal and substantial amount is required from the National Electronics Research Council which itself is dependent upon the generosity of industrial and other subscribers, including the I.E.R.E.

By succeeding in launching its first major research project N.E.R.C. will have justified its foundation, and in time will earn the gratitude of all engineers who wish to see technical progress unhampered by delay in securing information.

G. D. C.

* See the Presidential Address of Admiral of the Fleet the Earl Mountbatten of Burma, K.G. (*J. Brit. I.R.E.*, 6, Jan. 1946), and his paper read before the Royal Institution on 28th January 1966 (*The Radio & Electronic Engineer*, 31, April 1966).

INSTITUTION NOTICES

Symposium on ESRO 2

The Symposium organized by the Astronautics and Guided Flight Section of the Royal Aeronautical Society on 'A Progress Report on the ESRO 2 Programme', to which members of this Institution have been invited by the Society, has been postponed from 2nd November to Monday, 14th November.

The programme is as follows:

10.00–11.15 'Overall Project Management and Technical Progress' by C. R. Hume, M.B.E. (Associate Member) (Hawker-Siddeley Dynamics), A. Kutzer (ESRO) and H. L. Eaker (N.A.S.A.) (Chairman: W. N. Neat, F.R.Ae.S.)

11.30–12.45 'Systems Development' by P. Quetard (Engins MATRA). (Chairman: Dr. G. H. Hough, F.R.Ae.S.)

2.15–3.20 'Space Simulation' by S. M. Entres (ESTEC). (Chairman: L. H. Bedford, C.B.E., F.R.Ae.S. (Member))

3.45–5.15 'Scientific Experiments' by D. E. Page (ESLAB) (Chairman: I. Maddock, O.B.E. (Member))

5.15–6.00 General Discussion.

The Symposium will be held in the Society's Lecture Theatre at 4 Hamilton Place, London, W.1, and members wishing to take part should apply to the Sections Secretary of the Royal Aeronautical Society at that address for tickets.

Proceedings of Recent Conferences

Preprint volumes for the two I.E.R.E. Conferences held this year are still available and may be ordered from the I.E.R.E. Publications Department 8–9 Bedford Square, London, W.C.1.

These Conferences are:

'Applications of Thin Films in Electronic Engineering', held in London, 11th–14th July.
'Electronic Engineering in Oceanography', held in Southampton, 12th–15th September.

The charge for the volumes is in each case £6 postage paid and includes the supplement volumes to be published towards the end of the year, which will contain papers not preprinted, addresses and discussion reports.

Orders may now be placed by libraries and others preferring to receive their *Proceedings* of either of these Conferences within a single cover. These final *Proceedings* will be ready towards the end of the year, and will cost £6 each postage paid.

Conference on M.F., L.F. and V.L.F. Radio Propagation

A Joint Conference on the above subject is being sponsored by the Electronics Division of the Institution of Electrical Engineers (through its Professional Groups on Electromagnetic-Wave-Propagation, Television and Sound Broadcasting, and Radio Navigation and Location) and the Institution of Electronic and Radio Engineers (through its Radar and Navigational Aids Group). The Conference will be held at the Institution of Electrical Engineers, Savoy Place, London, W.C.2, from the 8th to 10th November 1967, and its scope will include:

Navigational Aids; Communication
The Effect of Ground, Coastline and Climate
Ionosphere—Irregularities, Absorption, Movement and Non-linear Effects
Waveguide Mode and Ray Theories
Excitation Factors
Phase and Field Strength Stability
Atmospherics
V.L.F. in the Magnetosphere
Sub-surface Propagation

The Organizing Committee, on which the I.E.R.E. is represented by Mr. C. Powell (Member), invite the submission of contributions of up to 2000 words for consideration for inclusion in the Conference programme. The texts of contributions should be submitted before 30th June 1967, but in the first place intending contributors should submit a synopsis (approximately 200 words) by 31st January 1967.

Further details of the Conference will be announced in due course.

The Institution Tie

Increasing costs of manufacture have made it necessary to increase the cost of Institution Silk Ties to 27s. each (post free). The cost of the Terylene Ties remains unchanged at 21s. each (post free).

The Institution Tie may only be obtained from the I.E.R.E. at 8–9 Bedford Square, London, W.C.1, and its wearing is restricted to members of the Institution (including registered Students).

Back Copies of the *Journal*

With reference to the request published in the September issue for back copies of *The Radio and Electronic Engineer*, a sufficient quantity has now been received from members and *no further copies of any issues are at present required.*

U.D.C.621.395.625.2: 681.844

The Floating Transcription Arm: A New Approach to Accurate Tracking with Very Low Side-thrust

By

A. R. RANGABE, B.A.†

Presented at a meeting of the Electro-Acoustics Group in London on 9th February 1966.

Summary: The limitations of existing pivoted arms are outlined, followed by brief details of some past and present radial tracking arms. The advantages of very low tip mass cartridges tracking at $<0{\cdot}2$ g, and the design requirements of suitable arms, are discussed as an introduction to the principles and design of a radial tracking arm with a side thrust of <3 dynes at $33\frac{1}{3}$ rev/min ('Trutrack'), in which the weight of the arm, head and counterweight is supported by a float immersed in silicone fluid.

Details of setting-up techniques and methods used to seal and package the liquid filled trough to withstand transport shocks of 100 *g* are given. The calculation and measurement of inertial forces resulting from eccentric and warped records, wow, frequency and damping of various vibrational modes of the arm are described. Details of the magnetic repulsion bearing system used in the device, with some performance curves, are given, and the problems of designing reliable low drag liquid signal contacts discussed. Materials and techniques used to ensure long term reliability of the device are described.

1. Introduction

1.1. *Pivoted Arms*

By far the most popular method evolved for tracking a cartridge across a record is the offset-head pivoted arm. This is basically a very simple system and a large number of variations on the theme have appeared over the years. Modern top quality pivoted transcription arms have reached a high degree of perfection. They are beautifully engineered precision instruments, which are more than adequate in meeting the demands of the lowest tracking weight cartridges available at present, tracking at about 1 gramme. However, should very much lower tip-mass cartridges, capable of tracking at significantly lower weights, become available, certain features inherent in even the best conventional pivoted arms would produce an undesirably high side thrust and variation in tracking weight. These features are as follows:

(i) horizontal and vertical pivot frictions,

(ii) forces due to lead stiffness,

(iii) inward force due to head offset.

This last force can be reduced to a great extent by a suitably arranged compensating weight or (in one case) magnets. However, complete cancellation of this force is not possible as there is a considerable variation in the coefficient of friction from record to record, and even within a record. This is shown in Table 2 which gives values of the amount of drag for a number of random records of various makes which have been measured by the author with a specially designed 'friction balance'. It is thought that it would be very difficult to produce a commercial pivoted arm in which the total side-thrust stemming from all the above causes could be held *consistently*, over long periods, to less than 30 dynes. Variation in tracking force on warped records due to pivot friction and lead stiffness would be rather less, but of the same order as above.

1.2. *Radial Tracking Arms*

In correctly-designed pivoted arms of average length there occurs a maximum horizontal tracking error of about $1\frac{1}{2}$ deg. This only produces a minor contribution to harmonic and intermodulation distortion.[1, 2] Nevertheless, over the years a number of radial tracking arms have been proposed or developed. The object of these devices is to make the reproducing stylus follow the same path across the record as that tracked by the cutting stylus, thus eliminating the aforementioned horizontal tracking error. To name a few of these systems, one of the earliest was a fearsome machine patented by G. H. Balmain in 1922, in which 'a carriage carrying the sound box and the horn . . . is carried by balls, rollers or wheels on rails above or laterally off the record so that the needle is guided to move directly towards the centre of the record'. Alternatively 'the rollers and like and the rails may be replaced by a bath of mercury'.[3] A recent version of this theme was the radial tracking arm developed by the B.B.C., for their own use.[4] In this an arm plus cartridge was fixed to a carriage

† Consultant, Stoneacre, Denmead, Portsmouth, Hampshire.

provided with rollers which travelled in a guide channel. The Mackie parallel tracking arm, which was available commercially in the 1950's, worked on a similar principle.

Unfortunately, a heavy price has to be paid with all rail guided devices for a small improvement in horizontal tracking angle. If there is any record eccentricity, the whole mass of the head, plus arm, carriage and counterweight assembly, has to be accelerated linearly resulting in much larger inertial forces at the stylus than those produced by a pivoted arm of the same total mass. Also, as there is no mechanical advantage in the system, the full bearing friction forces appear at the stylus and any dust or very small imperfections in the track or rollers, will contribute large additional forces. Furthermore, the full effects of signal lead stiffness will be felt by the stylus. Finally, to minimize side-thrust stemming from the effects of gravity, the guide rails have to be levelled to better than 1 minute of arc.

To complete the radial tracking arm story, mention must be made of two other quite different lines of approach to the problem, which are available at present: (a) the BJ 'Tan' arm, which works on the pantograph principle and is designed for reduced tracking error, (b) the ingenious Worden arm in which tracking error is eliminated by using an articulated head shell. Both these devices are essentially pivoted arms, and so do not suffer from the shortcomings of rail guides systems. Unfortunately, however, they still have the problems of normal pivoted arms, i.e. bearing frictions, lead stiffness, and variable side-thrust, which cannot easily be compensated.

1.3. *The Cartridge Problem*

Recent developments in bi-radial styli, standardization of vertical tracking angle, and stylus mass reduction, have resulted in a number of top grade cartridges capable of a very high standard of performance. Nevertheless, in striving for the ultimate in record reproduction there is still quite a long way to go.

In his paper entitled 'The rational design of phonograph pickups' Hunt[5] discusses a number of features in the design of cartridges which could be improved and shows that it is desirable to arrange for the resonance of the first compliantly supported mode of the stylus to be *at least* an octave above the highest recorded frequency. As far as the author is aware there is no existing top quality cartridge which meets this specification. These remarks apply to 0·0005 in radius diamonds. The introduction of bi-radial diamonds with an even smaller area of contact and hence a higher compliance between the tip and the record material reduces the frequency of this resonance still further.

In order to achieve this requirement, a further major reduction in effective stylus mass will have to be made. This will of course allow a corresponding reduction in tracking weight, which is highly desirable.

In another section of his paper Hunt also shows that a reduction in tracking weight results in a decrease in translation loss. All his results, including those based on Miller's analysis which he quotes, assume that the record material is not being stressed beyond the elastic limit. In fact, from some recent work by Barlow[6] there is good reason to believe that this is far from being the case, and hence the position is worse in practice than Hunt's idealized system predicts. In fact, according to Barlow's results, a cartridge tracking at 1 gramme with a 0·0005 in radius stylus tip is working in the fully plastic range of the record material, i.e. above the elastic limit, by almost two orders of magnitude. For a $0{\cdot}0008 \times 0{\cdot}0002$ in bi-radial diamond the situation is even worse, probably by a factor of two.

Actually, the position is not quite as gloomy as the above data would seem to imply, because up to about one order above the elastic limit, plastic flow only takes place *below* the surface of the record material, and only increases very slowly on repeated playing. Under these conditions, surface deformation is barely perceptible. Nevertheless, even to satisfy this second requirement, the tracking weight, for all deformation to be sub-surface on repeated playings, would have to be limited to about 0·2 g for a 0·0005 in radius diamond, and not much in excess of 0·1 g for a $0{\cdot}0008 \times 0{\cdot}0002$ in diamond. To track safely a maximum recorded acceleration of 1000 g means that using the values given by Hunt,[5] maximum effective stylus masses must not exceed 66 and 33 μg respectively, at the above tracking weights.

Using certain novel transduction principles, it is believed that there is a reasonable chance of producing cartridges to these specifications in the not too distant future. Unfortunately, when such heads become available, it is unlikely that they will perform satisfactorily, even with the best available conventional arms. The horizontal force of the order of 30 dynes at the stylus, referred to in the discussion on arms, would result in a highly asymmetric loading of the groove walls which is obviously most undesirable, and this coupled with the effects of horizontal pivot friction would probably result in groove jumping with moderately warped records.

2. A New Approach to the Tracking Problem

2.1. *Principle of the 'Trutrack' Arm*

In an attempt to overcome these problems, a study of possible alternative tracking systems was undertaken in the latter part of 1962. This eventually resulted in the design and development of a radial

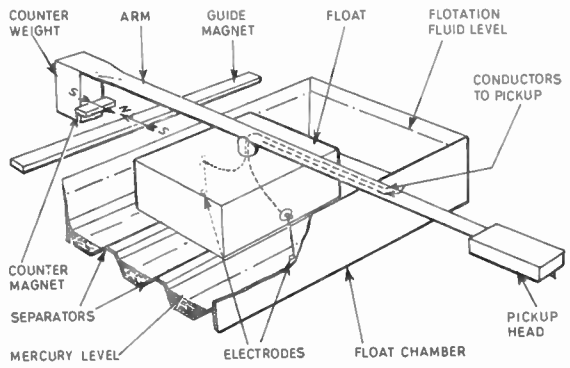

Fig. 1. A simplified schematic of 'Trutrack' arm.

tracking arm capable of tracking a concentric 33⅓ rev/min record with a side-thrust of less than 2 dynes at the stylus, and maximum horizontal tracking error of less than 0·5 deg. This arm has been called 'Trutrack' and will be referred to as such in this paper. The principles of this device, although simple, are so completely different from those of any existing or past tracking arm, that a fair amount of explanation is necessary. A highly simplified schematic drawing is shown in Fig. 1 in which only components essential to an understanding of the principles of the device have been included.

The fundamental differences between 'Trutrack' and existing arms come under three headings:

(a) The weight of the head, counterweight and arm is supported by a float immersed in a silicone fluid. The only resistance to motion through this fluid is that due to viscous forces, which at the tracking speed of the stylus are very low, amounting to an average side-thrust of the order of 1¾ dynes for a 33⅓ rev/min record.

(b) The pick-up output is transferred to the amplifier via 'liquid contacts'. The float chamber of the main trough is sub-divided by insulating separators into three shallow sub-troughs which contain mercury or a highly conducting electrolyte which is denser than, and will not mix with, the highly insulating silicone. A contact strip is fixed to the floor of each sub-trough, and wire contacts protrude into the conducting liquid from the underside of the float; signals are transferred from these to the fixed contacts via the liquid. The extra drag produced by these wire contacts moving through the liquid is negligible, and thus there are no forces resulting from lead stiffness and their attendant problems which arise in conventional pick-ups.

(c) To resist the forces due to frictional drag between the stylus and the record surface, a frictionless linear bearing has been developed, consisting of a system of permanent magnets. In its simplest form (see Fig. 1) it consists of a long guide magnet attached to the float chamber and a small magnet forming part of the counterweight assembly on the arm. These are both magnetized along their minor axes and aligned with like poles facing, and hence repelling, each other. With this arrangement, the record drag forces are balanced by the magnetic repulsion forces, and the head is made to follow a truly radial (parallel) track on the record.

The bearing system used has a very important advantage in relation to inertial forces on the stylus. Any horizontal movement of the stylus due, for instance, to record eccentricity, results in the arm-float-counterweight system *pivoting* in the horizontal plane, approximately about the centre of inertia, which in practice is near the centre of the float. As a result, this 'single pivot' radial tracking system gives the best of both worlds, i.e. accurate radial tracking at constant velocity and, at the same time, the reduced inertial forces of a pivoted arm. This, of course, is only made possible by virtue of the very low and constant tracking drag of the system. Another useful feature is that due to the very low natural frequency of vibration of the float assembly in the liquid, the arm assembly is virtually isolated from the motor board at the rumble frequency, and in some cases this results in a reduction of rumble.

2.2. *Description of Production Arm*

The general-purpose 'Trutrack' has been engineered to accept the majority of *existing* high quality cartridges, and can be adjusted to track from 0·5 g to 4 g. For reasons which will be discussed later, the shell arm counterweight assembly of this unit is not suitable as it stands for lower tracking weights. For this purpose an integrated head-arm-counterweight assembly will be produced when suitable ultra-low tracking weight cartridges become available. Brief details of this are given later in the paper. This can be easily substituted for the general-purpose arm, all other parts of the system remaining the same. The description which now follows refers to the general-purpose version.

A cut-away view of a complete 'Trutrack' unit is shown in Fig. 2, while Fig. 3 is a transverse section. A number of features already described in connection with the schematic drawing will be recognized. These are the plastic trough (1), liquid contacts (5), flotation liquid (4), stainless steel float (6), float contacts (23), fixed contacts (22), and guide magnet (14). Additional features and method of operation of a practical device are as follows. The trough (1) is fitted with a die-cast lid (2) which has a rubber O ring (24) bonded to its periphery. This is compressed into a 'V' groove in the edge of the trough wall by clamping bands (28).

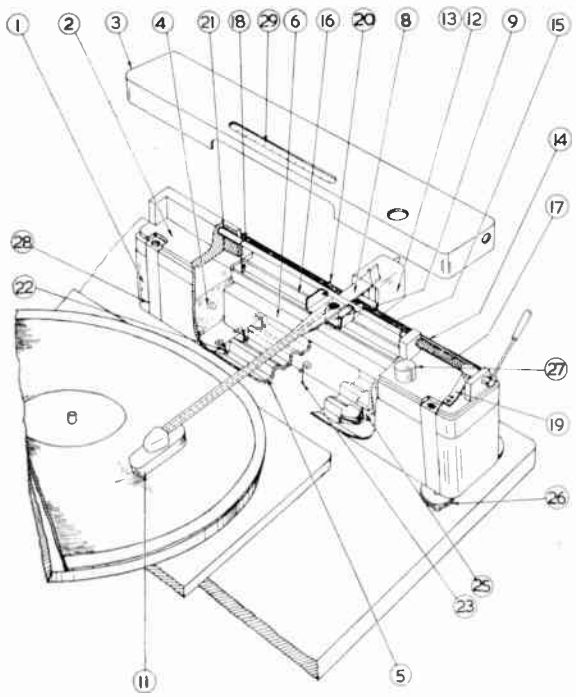

Fig. 2. A cut-away view of a complete 'Trutrack' unit.

The lid has an elongated slot through which extends a spigot (9) which is rigidly attached to the float at its lower end and the flattened portion of the arm (8) at the top. A rubber gasket (10) is recessed into the underside of the lid. The float is provided with a raised bead (7) round the top edge (see Fig. 3). The float can be raised via the arm by a mechanism described below until the bead makes contact with the rubber gasket. Under these conditions, with the float centred, it provides a dust-tight seal when the arm is not in use. A transit clamp is provided which applies a much higher pressure to this seal, making it possible for the device with the liquids in it to withstand shocks of up to 100 *g* in transit without damage or spillage.

The raise-and-slow-lower mechanism is very simple. A longitudinal stainless steel shaft (15) is located in low-friction dry bearings attached to the lid. A 'gantry' (16) consisting of two non-ferrous strips brazed together is suspended by flexibles (18) from the shaft. In the playing position the gantry is housed within the thickness of the lid, with the flattened portion of the arm above and clear of it. A 'fork' (20) is attached to the arm assembly. To raise the arm the shaft is rotated in a clockwise direction which has the effect of raising the gantry on the windlass principle until it contacts the arm. Further upward movement results in (a) the head being lifted off the record, and (b) the fork engaging

Key to Figs. 2 and 3

1. Trough
2. Lid
3. Dust cover
4. Flotation liquid
5. Liquid contacts
6. Float
7. Float bead
8. Arm
9. Spigot
10. Rubber gasket
11. Shell
12. Counter-magnet
13. Counterweight
14. Guide magnet
15. Shaft
16. Gantry
17. Torsion spring
18. Flexibles
19. 'Click' stop
20. Fork
21. Dashpot
22. Fixed contacts
23. Float contacts
24. 'O' ring
25. Equalizing channel
26. Levelling screws
27. Bubble level
28. Clamping band
29. Scale

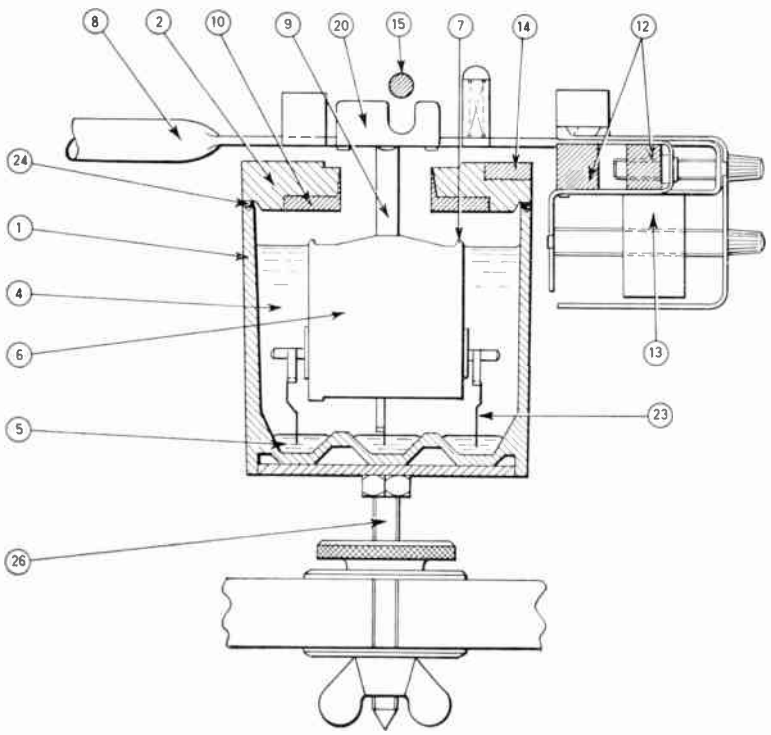

Fig. 3. A transverse section of 'Trutrack' arm.

the shaft, thus preventing rotation of the arm about the vertical axis. It can be retained in this position by a spring loaded 'click' stop (19). The arm can now be traversed by hand, and its position cued to within a few microgrooves with the help of the transparent scale (29) in the dust cover (3). Further rotation in a clockwise direction locks the float in the 'sealed' position as described above. To lower the arm from the cueing to the playing position, on any part of the record, the above process is reversed, allowing the windlass to unwind under the action of the torsion spring (17). The speed of rotation and hence of lowering of the head is controlled by a viscous dashpot arrangement (21).

The countermagnet assembly (12) is more elaborate than the single magnet arrangement shown in Fig. 1—a 'quadrupole' system is used (Fig. 5). This consists of a central attracting magnet, part of which is fixed and part adjustable, towards and away from the guide magnet by a calibrated micrometer screw. This attracting magnet is flanked by two fixed repelling magnets. The arrangement allows the correct radial tracking position to be set accurately over a range of ½ to 4 g. Apart from its range of adjustment this system provides an extremely 'stiff' magnetic bearing resulting in only very small excursions from the correct radial tracking position for large variations in record drag.

Below the countermagnet is the counterweight (13). This also has a calibrated micrometer screw adjustable from ½ to 4 g.

The arm consists of a very light high tensile aluminium alloy tube. Connections are carried through this on a flexible printed circuit. The front end of the tube is keyed to orient the adaptor, which is also locked in position by a taper pin at 90 deg to the key (Fig. 4). The adaptor carries an ultra-lightweight plastic shell (11) below it, which is merely a cover, *not* a structural member. Inside the shell there is a rectangular plastic 'biscuit' which is attached to the adaptor through the shell by three screws. The cartridge is in turn attached to the biscuit by the standard ½ in centre fixings, the whole thus forming an extremely rigid assembly.

A range of biscuits has been designed to compensate for height and horizontal stylus position variations between different cartridge types. Also, biscuits will be suitably 'slugged' to provide constant front-end weight, thus making it possible to set the tracking weight by the counterweight micrometer. By these means correct tracking position and weight is assured for any cartridge type.

There is a minor disadvantage in this system in that all cartridges have to be slugged to equal the heaviest. Of course, if preferred, full advantage can be taken of the exceptionally low front-end mass of the system (4 g including all fixings and connection hardware) by using a very light distance piece in place of the slugged biscuit and setting tracking weight with a stylus pressure gauge.

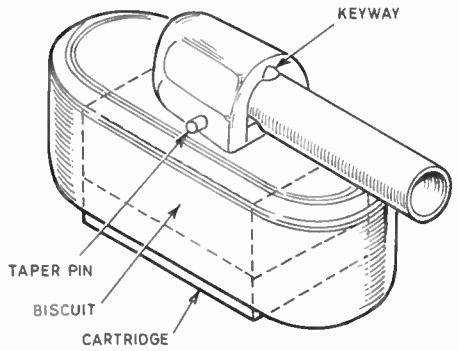

Fig. 4. Diagram showing the adaptor carrying the cartridge.

2.3. *Setting Up*

Although 'Trutrack' will operate within a reasonable degree of misalignment and out of level, certain alignment errors, discussed later, lead to an increase in horizontal tracking error, and hence side-thrust.

A fairly simple and foolproof alignment procedure has been devised which ensures that by carefully following the instructions, even an unskilled person can set up the unit correctly.

The alignment kit consists of:

(a) *An alignment protractor* in the form of a bent metal strip which can be clipped on to the unit. When this is slipped over the turntable spindle the unit is constrained so that the locus of the stylus is always along a radius of the turntable. To obviate the need for a drilling template the ends of the three levelling screws (26) are pointed, and the correct drilling positions are marked out by pressing them into the motorboard.

(b) *A 'U' shaped height gauge.* With the alignment protractor in position, the 'U' is dropped over the clamped arm where it extends over the turntable. Using the bubble (27) to keep it level the unit is lowered by means of the levelling screw jacks until the ends of the 'U' touch the turntable. This gives the correct height.

(c) *A plastic wedge* for mercury levelling. After transport the mercury may be unevenly distributed between the three sub-troughs. There is a raised equalizing channel connecting them at one end (25). By inserting the wedge under the single central levelling screw jack the trough is tilted sufficiently for the

mercury to join up across this channel. Thus, when the wedge is withdrawn, mercury is evenly distributed in the three sub-troughs.

This completes the setting-up procedure. Before removing the alignment protractor the unit is fixed in position by tightening butterfly nuts fitted to the levelling screws which project below the motor board. The unit has sufficient adjustment to accommodate a wide range of transcription turntables.

3. Performance Details

3.1. *Static Characteristics of the System*

3.1.1. Position of metacentre

The axis about which a floating object pivots, known as the metacentre, is given by

$$H = \frac{I}{V}$$

where H = height of metacentre above centre of buoyancy.

I = moment of inertia of waterline section about axis.

V = immersed volume.

In the case of a rectangular float, and for small deflections, this works out to

$$H = \frac{b^2}{12d}$$

where b = dimension perpendicular to pivoting axis.

d = immersed depth.

$0{\cdot}5\,d$ = centre of buoyancy.

Substituting appropriate values for the 'Trutrack' float gives:

$H = 0{\cdot}1$ in above centre of buoyancy (pivoting about major axis).

and $H = 3{\cdot}3$ in above centre of buoyancy (pivoting about minor axis).

The latter value means that the arm has considerable torsional stiffness pivoting about this axis as, for small angles, righting moment is proportional to the height of the metacentre above the centre of gravity, multiplied by the total weight and angle through which the float is displaced. In this case the metacentre is about 3 in above the centre of gravity.

3.1.2. Stylus-metacentre geometry

The design is such that the metacentre pivoting about the longitudinal axis is in the same plane as the record surface. This arrangement results in the minimum possible value of warp-wow.

3.1.3. Centre of gravity of head and counterweight relative to metacentre

For constant tracking force a line joining the centres of gravity of the counterweight and arm plus head assembly should pass through the metacentre, pivoting about the major axis. In the general purpose 'Trutrack' a compromise has been necessary as the result of limitations imposed by the adjustable counterweight/countermagnet assembly. This has resulted in the centre of gravity of the arm assembly being about 0·5 in above the metacentre. However, for the lowest tracking weight for which this arm is designed (0·5 g) the variation in weight does not exceed $\pm 8\%$ of the tracking weight for 0·25 in peak-to-peak warp.

In the special ultra-low tracking weight arm this problem does not arise as the centre of gravity is arranged to coincide with the metacentre.

3.1.4. Tracking errors

By using the alignment kit and built-in bubble level previously described the arm can be very accurately set up. However, the design is such that a reasonable amount of misalignment can be tolerated, without serious consequences. Thus, a longitudinal 'out-of-level' of the trough of up to $\pm 1\frac{1}{2}$ deg (bubble sensitivity is better than 10 min of arc) does not result in any increase in tracking error. A similar transverse tilt will produce a horizontal tracking error (due to tangential movement of the stylus) of just over $\frac{1}{2}$ deg in the worst case (inside grooves). There is no increase in tracking error for a longitudinal misalignment of $\pm\frac{3}{8}$ in in the horizontal plane. On the other hand, in the transverse direction, a departure of $\pm 0{\cdot}05$ in from the radial path results in a tracking error of ± 1 deg at the inside grooves.

3.2. *Dynamic Characteristics*

3.2.1. Viscous forces

The float, $6\frac{1}{4}$ in long by $1\frac{1}{8}$ in square section, is immersed to a depth of about 1·05 in in a dimethyl silicone fluid of 20 centistokes viscosity. This material has a very low temperature coefficient of viscosity $\left(0{\cdot}59 \text{ when specified as } 1 - \frac{v \text{ at } 210°\text{F}}{v \text{ at } 100°\text{F}}\right)$. The viscosity selected is a compromise between adequate damping of certain oscillatory modes of the arm to be discussed later, and reasonably low drag at the tracking velocity. Now, at the maximum velocity likely to be encountered in the system, the Reynolds number will be less than unity, hence the flow under all conditions is entirely laminar and Newtonian. Even so, precise calculation of viscous forces in various modes is extremely difficult and as a knowledge of these is required for drag and damping calculations, methods of measuring these directly were devised.

3.2.2. Linear tracking drag

When the float moves along the axis of the trough, the viscous forces resisting its motion consist of two parts: (i) the shear force arising from the movement of the submerged float surfaces past the stationary trough walls and, superimposed on this, (ii) the drag due to a counter-flow of fluid in the channels bounded by the float and trough walls. This results from the fact that fluid, being incompressible, is displaced from the forward to the after section of the trough, as the float moves. In practice, approximate calculations show that the forces due to (ii) are about an order up on those due to (i), and constitute the major part of the tracking drag.

An arrangement was set up to measure the linear drag forces accurately. This consists of an open trough filled with the liquid under test, with an appropriate float ballasted to float at the required depth. The float is provided with a 'towbar' in the form of a vertical wire cantilever. The float is towed along the trough at the required speed by connecting the top of the cantilever to a head tracking a record. The cantilever is appropriately calibrated in dynes thus making it possible to read drag directly. Results of tests with 20 cS silicone fluid give a drag of 1·75 dynes at 18°C+25% at 10°C and −20% at 30°C, at 33⅓ rev/min. For 45 rev/min the corresponding value at 18°C is 2·4 dynes. These drag values may be somewhat higher in practice, because of friction effects at the interface between the silicone and the mercury.

3.2.3. Mechanical resistance measurements

The values of the mechanical resistance of the float oscillating about (a) the vertical axis, and (b) the longitudinal horizontal axis, are required in connection with the calculation of forces at the stylus due to eccentric and warped records, as well as 'lead-out', and also for the determination of the damping factors of the effective head mass–stylus compliance resonances. The values of these mechanical resistances have been obtained by arranging for the float to oscillate in the silicone about the required axes at a sufficiently low frequency (~0·5 Hz) to make it possible to observe amplitudes of successive swings, and hence measure the decrement.

For oscillations about the vertical axis, the float, suitably mass loaded, is suspended in the trough by a wire, and made to execute torsional oscillations. The mechanical resistance is given by:

$$R_v = \frac{32dfI_v}{a^2}$$

where d = log decrement
f = frequency (Hz)
I_v = moment of inertia about vertical axis
a = length of float.

Substituting measured values, mechanical resistance $R_v = 1100$ dyne. s.cm^{-1}, in this mode.

For oscillations about the longitudinal axis, the float is provided with an arm ballasted with the required total weight distributed so as to make the arm stable and level when at rest. The arm is made to execute 'see-saw' oscillations about the metacentre and frequency and amplitude are measured as before.

In this case the mechanical resistance is given by:

$$R_h = \frac{8dfI_h}{b^2}$$

where I_h = moment of inertia about longitudinal axis
b = width of float.

Substituting numerical values, $R_h = 2100$ dyne. s.cm^{-1}.

3.2.4. Calculation of peak viscous forces

(a) *Record eccentricity.* Using the maximum eccentricity allowed in B.S. 1928,[7] the cumulative tolerances amount to 0·023 in. At 33⅓ rev/min this will produce a transverse peak stylus velocity of $V_e = 2{\cdot}1 \times 10^{-1}$ cm.s^{-1}. The peak force at the stylus will be given by:

$$F_{max} = \frac{V_e R_v a^2}{16l^2}$$

where l = distance of stylus to vertical pivot.
a = length of float.

Substituting $R_v = 1100$ dyne. s.cm^{-1}, $a = 6{\cdot}25$ in, $l = 7{\cdot}75$ in, $F_{max} = 9{\cdot}1$ dynes.

(b) *Lead out.* The maximum value for the lead-out velocity on a 33⅓ rev/min record is $V_1 = 4{\cdot}5 \times 10^{-1}$ cm.s^{-1}. (B.S. 1928.)

Substituting this value in the above expression gives

$$F_1 = 19{\cdot}5 \text{ dynes.}$$

3.2.5. Inertial forces

(a) *Centre of inertia.* As mentioned earlier, if the head is accelerated laterally, the arm-float-counterweight assembly pivots about the centre of inertia of the system. A knowledge of the position of this axis is needed before calculations of inertial forces at the stylus can be made. The position of the centre of inertia can be found by equating moments of the various components of the arm. For the general-purpose 'Trutrack', the axis about which the arm pivots, for transverse acceleration of the head, is very near the point of attachment of the arm, which is at the centre of the float.

(b) *Maximum inertial force at stylus due to record eccentricity.* Using the previous maximum eccentricity of 0·023 in and a total head mass of 16 grammes (Decca ffss. + adaptor) which is a maximum for

'Trutrack', the total effective mass about the vertical axis at the stylus is 22·5 g, including arm, float and counterweight contribution. The peak acceleration for eccentricity of 0·023 in at 33⅓ rev/min is approximately 7×10^{-4} *g*. This gives a force at the stylus of approximately 16 dynes peak.

It is interesting to note that even for this rather severe case, the maximum inertial forces are still at about half the value of the best that can be expected from pivot friction, etc., in a really good pivoted arm. It would appear from the above that with existing tracking weights far too much is made of the virtues of using ultra-light cartridges.

(c) *Maximum inertial force at stylus due to lead-out groove acceleration.* Hunt[5] quotes 0·015 *g* for 33⅓ rev/min and 0·026 *g* for 45 rev/min as the maximum lead-out groove accelerations likely to be encountered. With a 22·5 g effective head mass, these would result in a horizontal force at the stylus of 340 dynes and 520 dynes respectively. Thus a system of this effective head mass, tracking at 0·5 g, would have a reasonable factor of safety at 33⅓ rev/min (including the small additional contribution from viscous drag, see Sect. 3.2.3), but it would be very marginal at 45 rev/min. It would obviously be quite unsuitable for much lower tracking weights.

3.2.6. Resonances and damping

(a) *Effective head-mass-stylus compliance resonance.* The resonant frequency in the vertical mode will depend on the vertical compliance of the cartridge, assuming the effective head mass is kept constant. With a constant effective head mass of 20·5 g about the longitudinal axis it will range from 7·5 Hz for a compliance of 25×10^{-6} cm dyne^{-1} to 16·5 Hz for a compliance of 5×10^{-6} cm dyne^{-1}. To find Q in these two cases the value of the viscous resistance for the float pivoting about the longitudinal axis is used, assuming the arm-float assembly to be infinitely rigid. The effective damping at the stylus turns out to be very small indeed, $Q > 100$ in both cases, due to the highly adverse mechanical advantage (from this point of view) of the arm to half-width of float lever ratio, which appears squared in the damping equation.

As in practice it is known that the arm appears to be reasonably well damped in this mode, some tests were carried out to find out what was happening. Firstly the arm was set up with the correct mass on the front end to which was firmly attached a coiled spring of the required compliance. The lower end of the spring was driven by a variable frequency mechanical vibrator with a peak-to-peak amplitude of about 0·004 in. Resonance magnification was measured by means of an 'optical triangle' attached to the weight. For a resonant frequency of 15 Hz the Q was about 15. Further tests were carried out using a vertically modulated sliding frequency test record in conjunction with a Decca FFSS Mk. II head. Unfortunately, due to low recorded velocity and rather high rumble and hum level in the equipment it was not possible to obtain an actual figure for Q, but it can be stated that this was not greater than 5, which is acceptable.

There are some tentative theories to account for these wide discrepancies, but a complete explanation will have to await further tests. No damping measurements have been made for the transverse mode. No trouble has ever been experienced from this resonance.

(b) *Arm resonances.* Resonance searches have been made using B and K Stereo Gliding Frequency Test Record QR 2009 (20–20 000 Hz) and test record TCS 102, with a variety of cartridges. No resonances which could be attributed to the arm were observed in any of these tests.

3.2.7. Resonances specific to 'Trutrack'

It is obvious that with a system with so many degrees of freedom there are a number of additional resonances to be considered which could affect the performance of the arm.

(a) *Transverse oscillations of float-arm assembly in trough.* Two types of transverse oscillations are possible:

Type I. Float-arm mass-magnetic 'compliance' resonance. If undamped this *could* be excited by variations in record drag. Fortunately, as its natural frequency is well below 1 Hz, and because of the combination of the viscous resistance of the silicone in the transverse direction, added to the Coulomb damping arising from record friction, there is considerable over-damping of this resonance at all tracking weights. Should longitudinal movement of the stylus occur, due to variation in record drag, it is very small (Section 3.3.2) and takes several seconds to reach equilibrium. There is no measurable wow produced.

Type II. Float-arm mass plus liquid moving transversely relative to trough. This has been christened the 'slop' resonance. It is excited if the whole motor-board is jolted or shaken, thus setting up a transverse oscillation of the liquid in the trough. In this mode the float is carried with the liquid and forces the stylus to oscillate tangentially, producing wow at a frequency of about 2 Hz, slightly above critical damping. In other words, a transverse jolt will produce a total of two or three wows in a period of about a second. Fortunately, quite a large jolt is required to produce a noticeable effect, and under average domestic conditions with the equipment housed in a reasonably solid cabinet, no trouble should be experienced.

Table 1

Comparative performances of the general purpose and ultra light integrated arms

Parameter†	General purpose arm	Ultra-light integrated arm
	STATIC CHARACTERISTICS	
Metacentre (Sect. 3.1.1)	Major axis 0·1 in above c. of b. Minor axis 3·3 in above c. of b.	Major axis at c. of b. approx. Minor axis >3 in above c. of b.
Stylus-metacentre (Sect. 3.1.2)	In same plane	In same plane
Centre of gravity (Sect. 3.1.3)	Head–counterweight axis 0·5 in above metacentre	Head–counterweight axis on metacentre
Alignment tracking errors (Sect. 3.1.4)	(1) None for $\pm 1\frac{1}{2}^\circ$ longitudinal tilt (2) $\pm\frac{1}{2}^\circ$ for $\pm 1\frac{1}{2}^\circ$ transverse tilt (3) None for $\pm\frac{3}{8}$ in longitudinal position error (4) $\pm 1^\circ$ for $\pm$ 0·05 in transverse position error	As general-purpose arm
	DYNAMIC CHARACTERISTICS	
Viscous forces:		
Linear tracking drag (Sect. 3.2.2)	1·75 dynes at 18°C +25% at 10°C −20% at 30°C	Approx. 1 dyne at 18°C (Wetted surface approx. 50% of general-purpose arm)
Record eccentricity (Sect. 3.2.4(a))	9 dynes (0·023 in max. $33\frac{1}{3}$ rev/min B.S. 1928: 1961)	Approx. 4·5 dynes
Lead out (Sect. 3.2.4(b))	19·5 dynes ($33\frac{1}{3}$ rev/min. B.S. 1928: 1961)	Approx. 10 dynes
Inertial forces:		
Record eccentricity (Sect. 3.2.5(b))	16 dynes peak 22·5 g effective head mass (0·023 in max. $33\frac{1}{3}$ rev/min. B.S. 1928: 1961)	3·3 dynes 4·6 g effective head mass
Lead out (Sect. 3.2.5(c))	340 dynes 22·5 g effective head mass 0·015 *g* (Hunt). $33\frac{1}{3}$ rev/min	70 dynes 4·6 g effective head mass
Resonance and damping:		
Vertical–head mass-stylus compliance (Sect. 3.2.6(a))	7·5 Hz at 25×10^{-6} cm/dyne 16·5 Hz at 5×10^{-6} cm/dyne 20·5 g effective head mass $Q \simeq 5$	9·5 Hz at 100×10^{-6} cm/dyne 13·5 Hz at 50×10^{-6} cm/dyne 2·8 g effective head mass $Q \simeq 2$ (using integrated damped resonant absorber)
Horizontal	No measurements	—
Arm resonances (Sect. 3.2.6(b))	None observed 20–20 000 Hz	As general-purpose arm
Resonances specific to 'Trutrack'		
Transverse oscillations of float-arm (Sect. 3.2.7(a))	Frequency <1 Hz Considerably overdamped $Q<0\cdot5$	As general-purpose arm
Longitudinal oscillations of float-arm (Sect. 3.2.7 (b))	As above	As general-purpose arm
Vertical isolation (Sect. 3.2.7 (c))	22dB	—
	MAGNETIC BEARING PERFORMANCE	
Tracking accuracy (Sect. 3.3.2)	$\pm\frac{1}{4}^\circ$ (3·25 g tracking weight inside grooves) for max. variation in mean record drag	Better than general-purpose arm

† The sub-sections indicated discuss the parameters.

(b) *Longitudinal oscillations of float-arm assembly in trough.* Here again, as in the transverse case, two types of oscillation are possible. In this case, for the first type, the lateral magnetic compliance is very high, hence the frequency of oscillation is much lower than the previous case (0·07 Hz). As a result the oscillation is highly overdamped by the silicone and cannot cause any trouble. The longitudinal 'slop' mode is under-damped as in the previous case, but on this occasion all that happens is that the float pivots a degree or so about the stylus, resulting mainly in one or two cycles of increased tracking error. In any case the previous remarks about excitation apply here.

(c) *Isolation of vertical vibration.* The natural frequency of the float-arm assembly moving vertically is determined by the total mass, and the buoyancy compliance, and works out at about 3·3 Hz. Using 1100 dyne. s.cm^{-1} the value of mechanical resistance for the float oscillating about the vertical axis (Section 3.2.3) which should be of the same order of magnitude. The isolation factor against the 22 Hz rumble frequency vertical component is found to be of the order of 22 dB. There will be a measure of isolation against the horizontal rumble components by the 'slop' modes but they have not been measured or calculated yet.

Table 1 gives a résumé of the main performance details above, together with the corresponding values for the proposed ultra-light-weight integrated arm.

3.3. *Magnetic Bearing Performance*

3.3.1. Specification of requirements and development of the quadrupole system

The magnetic repulsion linear bearing used in 'Trutrack' has presented some of the most difficult design problems encountered during the development of the device. The ideal specification and design tolerances required for magnetic material to produce a bearing with the required tracking accuracy, is as follows:

(1) High coercivity and outstanding stability in the presence of varying demagnetizing fields.

(2) Highest possible slope of force/distance curve.

(3) Extremely uniform field over central 5 in for 10 in long guide magnet.

(4) High repulsion force per unit area.

(5) Good dimensional accuracy and straightness (for long guide magnet).

(6) Wide tolerance to vertical misalignment of countermagnet.

(7) Simple method of adjustment to compensate changes in tracking weight.

(8) Uniformity and reproducibility of production batches of material.

(9) Reasonable cost.

Unfortunately, a number of these requirements are incompatible in practice with available materials. McCaig[8] points out that for repulsion systems in which like poles are in close proximity, or may even touch, materials with a high coercive force are required. The readily available barium ferrite ceramic materials have suitable properties for many repulsion applications. For the 'Trutrack' bearing this type of material would satisfy requirements (1), (2) and (4) admirably. On the other hand, there is some doubt as to whether requirements (3), (5) and (8) could be met without pre-selection and possibly an expensive grinding operation, which would prejudice requirement (9). Another difficulty is that a ceramic guide magnet of the required cross-section would be very slender and fragile, and would lead to breakages in assembly.

Fortunately, another type of material with a reasonably high coercivity known as 'bonded Feroba' is available. This is still basically barium ferrite, but in this case it is in the form of a powder which is intimately mixed with a plasticized p.v.c. binder and extruded in continuous lengths in a variety of sections. Being extruded, the strip has a very uniform section, also it is flexible and can be jigged to the required linear accuracy. It has adequate coercive force (1150 Oe) and stability, but as the amount of magnetic material per unit volume is much less than that of the solid ceramic, BH_{max} is only 0·4 megagaussoersted compared with values of 1·0 and 2·7 megagaussoersted for ceramic Feroba I and III.

In terms of the specification above this material is reasonably adequate for requirements (1), (3), (4), (5), (8) and (9), but is disappointing in its force vs. distance performance (2), which of course is vitally important, as it determines the transverse stiffness of the magnetic bearing, which is required to be as high as possible.

These remarks apply to a simple dipole system using bonded Feroba for both magnets, as shown in Fig. 4. Other configurations, using N and S poles adjacent, facing similar poles, were tried, but although some of these arrangements were better, they invariably failed to meet requirement (6) of the specification, and also took up too much space vertically.

Eventually the arrangement which has been named the 'quadrupole' system was devised. This was described briefly in Section 2.2. This arrangement has a number of advantages; not only does it make a reasonably stiff magnetic bearing possible, using bonded Feroba, but by suitable design it also provides means for varying the repulsion force to balance the drag over a tracking weight range from 0·5 to 4 g. Details of the arrangement are shown in Figs. 3 and 5. In the plan view, the two-part attracting magnet,

flanked by the repelling magnets, is shown. The rear portion of the attracting magnet is in contact with the front portion for minimum force and can be moved back by the screw until it is $\frac{5}{8}$ in clear for maximum. Note that the repelling magnets are slightly offset to compensate for the small tracking error resulting from viscous drag on the float at tracking velocity. In the elevation view it will be seen that the counter-magnets extend above and below the guide magnet which allows a reasonable latitude in longitudinal out-of-level of the trough without appreciable change in the magnet gap. This meets requirement (6).

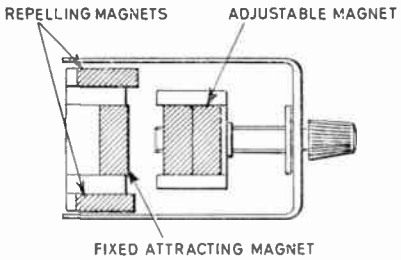

Fig. 5. Details of the 'quadrupole' system of magnetic bearing.

The performance of the 'quad' system is shown in Fig. 6 for three different settings of the adjusting magnet. A dipole equivalent to the high drag 'quad' setting is also shown for comparison. Re-plotting these results on log-log paper, reveals some interesting differences. Whereas at large distances the 3·25 g dipole obeys a cube law (which it should do!) it falls to a 1·1 exponent at the working gap. The equivalent numbers for the 'quad' at this setting are 3 and 1·75; at the 2 g setting 4 and 2·7, and at the 0·5 g setting 3·3 and 2·7. From this it is apparent that the 'quad' system provides a much 'stiffer' bearing than the dipole, with the added advantage that it can be made easily adjustable without much change in stiffness over a large part of the range.

3.3.2. Tracking accuracy

Both small radial and tangential tracking errors will result in an increase in side-thrust equal to the drag multiplied by the angular error. The 'quad' system makes much greater demands on the uniformity required of the guide magnet field than the simple dipole. Nevertheless, by using specially developed techniques for magnetizing the guide magnet, the tracking error in the worst case (0·5 g tracking) is $< \pm \frac{1}{2}$ deg, as magnetized, i.e. $< \pm 1$ dyne additional side-thrust at this tracking weight. In any case this error can easily be reduced as required by the judicious application of a few small nickel corrector strips across the guide magnet during initial testing. Thereafter, the stability of the bonded Feroba material appears to be adequate, no significant variations in performance occurring during life.

Using Table 2 in conjunction with Fig. 6, an approximate assessment of tangential tracking error can be made. Variations of drag within a record occur randomly, and in general peak values seem to be of short duration. Thus, due to the overdamping of the arm resonance (Section 3.2.6), in practice there is substantially no movement about the mean value. This only applies if the record is scrupulously clean. Sticky deposits and fingerprints may result in a very large increase in drag.

Table 2

Variations in drag of $33\frac{1}{3}$ rev/min long playing records of various makes

Decca ffss. Mark II, tracking at 3 g, 0·0005 in radius stylus

Record Make and number	Drag in dynes Minimum	Maximum
D.G.G. SPLM 138777(2)	520	700
Capitol SP 8518(1)	730	880
H.M.V. SXLP 20038(1)	660	750
R.C.A. RB 16170(2)	600	710
Decca EXP-BS-223-IN (New unmodulated test record)	660	770

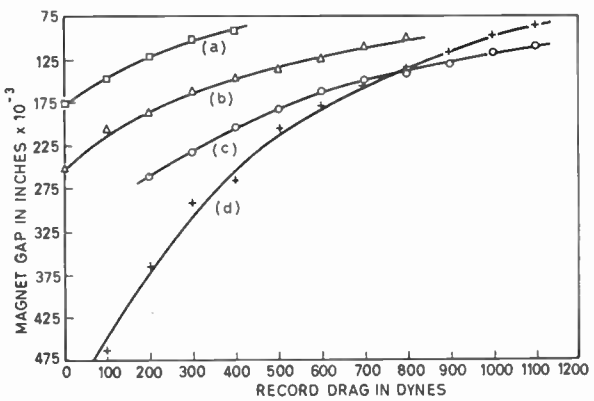

Fig. 6. Graph showing the performance of the 'quad' system for three different tracking weights. (a) 0·5 g, (b) 2·0 g, (c) 3·25 g, (d) dipole system for comparison.

Assuming that the spread in average drag between records shown in Table 2 is of the right order (605 to 810 dynes) the spread in tangential tracking error for the worst case (inside grooves, 3·25 g tracking weight) will be $< \pm \frac{1}{4}$ deg.

4. The Liquid Contact Problem

There are two types of liquid contact available for this application:

(1) metals liquid at room temperature, and

(2) electrolytes.

Of the limited available range of metals and alloys liquid at room temperature the only possible candidate for this application is mercury. On the other hand,there are a great number of electrolytes, mostly aqueous and non-aqueous solutions of neutral salts, which can be used.

4.1. *Specification of Liquid Contact Requirements*

In both systems the 'contact liquid' is covered by the 'flotation liquid'. (These are abbreviated to c.l. and f.l. respectively below.) The two have to be considered in conjunction. As in the case of the magnetic bearing, the combination has to meet a most exacting specification which is as follows:

(a) f.l. should have the highest possible density, and c.l. density should be higher still;

(b) c.l. should have adequate conductivity;

(c) f.l. should be a good insulator;

(d) f.l. should have a low surface tension;

(e) two liquids should be mutually compatible, and should not react chemically;

(f) both liquids should be compatible and non-reactive with other materials in the system in contact with them;

(g) both liquids should be stable over a wide range of ambient temperatures for very long periods;

(h) vapour pressure of both liquids should be low enough to ensure no appreciable loss by evaporation during life at the highest working temperature of the device;

(i) f.l. should have the appropriate viscosity with the lowest possible temperature coefficient;

(j) freezing point of both liquids should be below the lowest ambient operating temperature;

(k) cost of liquids should be low (of the order of a few shillings).

An initial evaluation was carried out of a number of possible flotation liquids and both mercury and a range of electrolytes for the contact liquid. A number of long chain esters and hydrocarbon oils were tried and also subjected to accelerated life tests, in conjunction with mercury and various electrolytes. In general, they failed sooner or later to meet conditions (e) or (f) or both, and left a lot to be desired as regards (i) and sometimes (a). The other possible liquid on the list was dimethyl silicone fluid. This appeared to meet every condition of the specification admirably, with the unfortunate exception of (k); the price was several times that of other liquids tested.

As regards the contact liquids, mercury was rejected during the initial evaluation as it posed some very difficult problems as regards purification, reliable contacts, and levelling.

4.2. *Silicone–Electrolyte System*

On the other hand there appeared to be a number of electrolytes which behaved well, although contact resistance (b), even with large area float contacts, was on the high side ($\sim 10\,\Omega$). After prolonged tests, including extended high temperature life tests, a system consisting of an aqueous electrolyte with a silicone fluid over it was selected and is described elsewhere.[9] About a month later assembly of a pre-production batch of units to this specification was started. A new batch of silicone fluid from the same source as before was used for these. Results were disastrous. The first unit put on test 'seized up' in about three days. Investigation showed the cause to be a layer of a highly viscous, jelly-like substance at the interface between the c.l. and f.l. The viscous drag on the float contact leads, which pass through this interfacial film, was sufficient to increase the drag on the float by 5 to 10 dynes. According to all the published information, and experts consulted at various times, there is no possibility of hydrolysis or other action between dimethyl silicone fluid and a neutral aqueous solution at room temperature and indeed nothing of the kind had been observed during previous tests. As a result a very extensive investigation was put in hand, everything *except* the silicone fluid being suspected. Finally, after about four months intensive testing and elimination it was established with almost complete certainty that the particular batch of silicone was to blame, probably due to insufficient removal of reactive chlorosilane intermediates. Thus, to be reasonably certain of avoiding trouble over a period of years, maximum impurity limits would have to be very rigidly specified and controlled.

4.3. *Silicone–Mercury System*

During the course of these investigations it was decided to take another look at the silicone–mercury system. This has a number of major problems:

(1) Even fairly pure mercury (<20 parts in 10^6 total impurities) 'tails' quite badly, due to a solid surface layer of oxidized base metal impurities, between the mercury and the silicone.

(2) Unless very fine (50 μm) wire contacts are used the drag due to this interfacial layer is unacceptably high.

(3) Using such fine non-amalgamated wires contact becomes unreliable, particularly when only a few millivolts are available to break down the insulating film which appears to form on the wire and/or mercury in silicone.

(4) The mercury breaks up into beads during transport. Due to the formation of the aforesaid film coupled with the high surface tension, these beads will frequently not join up during the levelling operation (Section 2.3 (c)), resulting in variations in level between the sub-troughs and discontinuous contact in these channels.

It was this formidable list of problems which initially prompted a decision in favour of the silicone–electrolyte system. In re-assessing the problem it was known that by increasing the area of contact between an unamalgamated wire and mercury sufficiently, adequate contact reliability was obtained. Thus, one line of attack was to produce sufficiently pure mercury to reduce the thickness of the surface oxide film, and hence the drag on quite a large diameter wire to an acceptably low value. A new method of refining the metal was developed. This made it possible to produce mercury so pure, that after several months in moist air, or under silicone, there was no 'tailing' and no visible film formed. Using this material the drag was very low indeed with wires large enough to give reliable contact. The drawback to this technique is that the system is extremely vulnerable to accidental contamination of the mercury during life, which makes it potentially unreliable. The problem of getting the mercury to join up under the silicone was finally solved by lining the sub-troughs with amalgamated pure silver foil. Mercury will dissolve 0·065 atomic per cent of silver at room temperature.[10] When this is saturated it can co-exist in contact with the silver–mercury alloy foil without further significant change in either component, and as silver does not oxidize under these conditions, no high drag interface is produced. When beads of mercury come into contact with this large surface area of mercury wetted alloy, they join up immediately and thus continuous liquid contacts in the sub-troughs are always assured.

The development of this stable silver–mercury alloy system also pointed the way to a much better solution to the float contact drag problem. A number of arrangements were tried, and finally a highly successful design was evolved. Briefly, this consists of a 0·002 in diameter hard drawn tungsten wire, with a 0·03 in square silver 'flag' swaged to its tip. In use, the amalgamated silver is immersed below the surface of the mercury and the non-amalgamating tungsten projects through the interface.

This system gives the best of both worlds: stable and reliable contact resistance from the silver–mercury contact and, the flag being completely immersed, it is only subjected to the *bulk* viscous drag of the mercury, which is very low indeed. On the other hand, the tungsten wire is fine enough to give an acceptably low interfacial drag even when laboratory grade mercury with up to 200 parts/10^6 of impurities is used. As this is about three orders higher than the impurity level of the mercury which will be used in practice, there is a reasonable factor of safety against accidental contamination during life.

Having finally succeeded in producing a reliable silicone–mercury system, it was decided to change over to it, as it has a number of advantages over the electrolyte system, as follows:

(1) Much lower contact resistance ($< 1\ \Omega$).

(2) Lower capacitance when fully screened (about 5 pF per channel).

(3) No loss of contact liquid by evaporation.

(4) A much higher factor of safety from interfacial contact drag.

Unfortunately, during the course of the development the cost of mercury soared up by a factor of four, which has added a significant amount to the production cost of the unit.

4.4. *Reliability and Material Compatibility Problems*

In such a complex system, with unusual combinations of materials and stringent performance specification, the utmost care has to be taken in the selection and specification of the various components. The silicone–electrolyte trouble is a very good example of this. As far as possible a 'belt and braces' philosophy has been adopted in the design. To cite a couple of examples: all rubbers are affected to some extent by silicone fluid. To make certain of reliable seals, a rubber with the highest silicone resistance is specified, and for good measure it is coated with a film of polythene, which is unaffected by the fluid. The second example relates to the float. This is a fabricated structure some parts of which are soldered. These soldered sections could come into contact with mercury during transport. As this would almost certainly result in a leaking float *and* badly contaminated mercury, an anodic passivation treatment has been developed for the solder. On a life test, this appears adequate protection against mercury, but to make doubly certain, the float is coated with a film of epoxy varnish.

No ferrous materials are used in components anywhere near the magnetic system, with the exception of austenitic stainless steel, to a rigorous specification, for the float. All highly stressed components are of high tensile beryllium copper, aluminium alloy and brass.

Finally, a question frequently asked is: what is the effect of dust on the performance of the system? Briefly, particles, or groups of particles, smaller than the shortest distance between fixed and moving parts (0·08 in) both in the liquid and outside, have no effect. Long filaments, of lint or hair, *can* upset tracking accuracy, if judiciously placed. However, the dust cover on the unit is designed to give sufficient protection.

5. Acknowledgments

Thanks are due to Mr. D. G. Jacquess, of the Decca Record Company Ltd., for the loan of special test records; to Garrard Engineering Limited for carrying out arm resonance investigations; to Mr. Rex Baldock for checking the manuscript and helping with the demonstration; and to Mr. R. Snell for setting up the demonstration.

6. References

1. H. G. Baerwald, 'Analytic treatment of tracking error', *J. Soc. Motion Picture Engrs*, **37**, p. 591, November 1941.
2. J. Walton, 'Pickup design', *Wireless World*, **66**, No. 8, p. 340, July 1960.
3. Brit. Pat. No. 177,215 (1922).
4. P. J. Guy, 'Disc Recording and Reproduction', Fig. 8.12, p. 112. (Focal Press, London, 1964).
5. F. V. Hunt, 'The rational design of phonograph pickups', *J. Audio Eng. Soc.*, **10**, No. 4, p. 274, October 1962.
6. D. A. Barlow, 'Groove deformation in gramophone records', *Wireless World*, **70**, No. 4, p. 160, April 1964.
7. B.S. 1928: 1961. 'Specification for Gramophone Records and Reproducing Equipment'.
8. M. McCaig, 'Permanent Magnets for Repulsion Devices', Permanent Magnet Assn., Tech. Bulletin, **1**, No. 4, 1965.
9. A. R. Rangabe, 'The Trutrack pick-up arm', *Hi-Fi News*, **8**, No. 11, April 1964.
10. L. Addicks, 'Silver in Industry', pp. 93–5 (Reinhold, New York, 1940).

Manuscript received by the Institution on 9th February 1966. (Paper No. 1069/EA32.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **September 1966**

September 1966	24-hour mean centred on 0300 U.T.			September 1966	24-hour mean centred on 0300 U.T.		
	GBZ 19·6 kHz	MSF 60 kHz	Droitwich 200 kHz		GBZ 19·6 kHz	MSF 60 kHz	Droitwich 200 kHz
1	− 301·0	− 301·3	− 0·4	16	− 300·8	− 299·9	+ 0·9
2	− 301·3	− 300·0	+ 0·1	17	− 302·5	− 299·8	+ 1·2
3	− 301·6	− 299·4	0	18	− 302·6	− 300·0	+ 1·1
4	− 300·0	− 299·4	0	19	− 300·8	− 300·1	+ 1·3
5	− 303·1	− 299·7	− 0·2	20	− 299·1	− 300·2	+ 0·8
6	− 302·7	− 300·2	− 0·2	21	− 301·0	− 300·2	+ 1·5
7	− 303·4	− 300·2	+ 0·2	22	− 301·0	− 300·4	+ 1·3
8	− 298·6	− 300·2	+ 0·6	23	− 299·8	− 300·7	+ 1·1
9	− 297·1	− 299·9	+ 0·7	24	− 296·2	− 299·9	+ 1·4
10	− 298·1	− 300·1	+ 0·8	25	− 297·8	− 300·0	+ 1·0
11	− 298·1	− 300·6	+ 0·5	26	− 299·6	− 300·0	+ 1·4
12	− 298·2	− 300·9	+ 0·1	27	− 300·2	− 300·2	+ 1·7
13	− 298·2	− 300·7	+ 0·2	28	− 298·0	− 300·4	+ 1·5
14	− 297·2	− 299·9	+ 0·4	29	− 298·6	− 300·3	+ 1·8
15	− 298·3	− 299·9	+ 0·6	30	− 299·3	− 300·5	+ 1·8

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

U.D.C. 557.570.535

Some Physical Aspects of the Ionosphere

By

Professor
D. A. BELL, M.A., B.Sc.,
Ph.D., C.Eng.†

Summary: The paper surveys the current state of knowledge of the properties of the ionosphere, in particular those which have become more important or more accurately known through the use of sounding rockets and artificial Earth satellites. The atmosphere and its electrical properties are first discussed. An account of wave-electron interaction and incoherent scattering from electrons is followed by consideration of the properties of the D, E and F regions, and finally scatter propagation is briefly discussed.

1. Introduction

From the point of view of radio communication, the ionosphere is that region of the Earth's atmosphere[1] in which the density of electrons is sufficient to cause reflections or attenuation of electromagnetic waves below, say, 50 MHz; and it is conventionally divided into four regions, D, E, F1 and F2. The characteristics of direct interest in each case are:

(1) The maximum electron density, N_m, which governs the highest frequency which can be reflected.

(2) The height, h_m at which N_m occurs, since this controls the geometry of the transmission path (single-hop or multi-hop).

(3) The collision frequency of electrons with other particles: the attenuation depends on the joint values of N and ν in all regions through which the wave has passed.

In addition to the major and reasonably predictable phenomena there are various 'sporadic' phenomena which can be used for communication with the aid of modern equipment.

The production of free electrons by the ionization of atoms or molecules requires energy which is supplied by various kinds of radiation, mostly of solar origin, which impinge on the earth's atmosphere; these include photons from x-ray to ultra-violet, protons, cosmic rays and the gross particles known as meteors. The numerical density at any point depends both on the rate of production of free electrons and on their rate of loss, the latter being influenced by recombination, diffusion and drift. Recombination and diffusion depend on the properties of the surrounding gas, but drift may be of electrodynamic origin.

2. The Earth's Atmosphere

A great deal of information about the ionosphere has been deduced from radio measurements, in particular from vertical-incidence soundings, but this has now been supplemented by data from instruments carried on rockets and satellites and from the perturbation of satellite orbits by atmospheric drag, particularly in relation to the particle density in interplanetary space and in the Earth's atmosphere above 600 km. Figure 1 shows a logarithmic plot of estimated atmospheric density versus height, based on various sources: the curve is of *density* rather than pressure because the experimental data are usually indicative of density and transformation to pressure involves the temperature which is a rather uncertain quantity.

The range from 0 to 20 km is derived from standard physical data (see for example, Reference 29); and the difference between summer and winter is negligible on the scale of Fig. 1. From 30 to 100 km the data came from Nicolet[2]; for 120 and 160 km from rocket data[29] and from 180 to 700 km from King-Hele's analysis[3] of the effect of atmospheric drag on satellite orbits. Gringauz[9] deduced from ion-probe measurements on *Sputnik III* (up to 1000 km) and *Lunik III* (up to 200 000 km) that O+ ions predominate up to 1000 km, but then protons predominate, falling to a numerical density of 1000 particles/cm^3 at 2000 km, 100 at 20000 km and 1·5 particles/cm^3 average in interplanetary space, though the latter rises to 25 particles/cm^3 in solar corpuscular streams. The scale in Fig. 1 has been made extraordinarily coarse in order to cover in one diagram the whole range of 19 decades of pressure. On this scale either of the points marked with crosses at 1600 km would be acceptable though in fact they represent contradictory hypotheses. The left-hand cross indicates the atmospheric pressure that would be necessary to account for the drag on the balloon satellite *Echo* by atmospheric friction only. The right-hand point arises from an interesting hypothesis of Drell *et al.*[12] that the drag is electrodynamic in origin. Since *Echo* is a conducting surface moving through the Earth's magnetic field, there will be an e.m.f. induced across it. But the surrounding medium is ionized and therefor conducting, so that with the aid of photo-electric emission to transfer electrons from the metal surface

† Department of Electronic Engineering, The University of Hull.

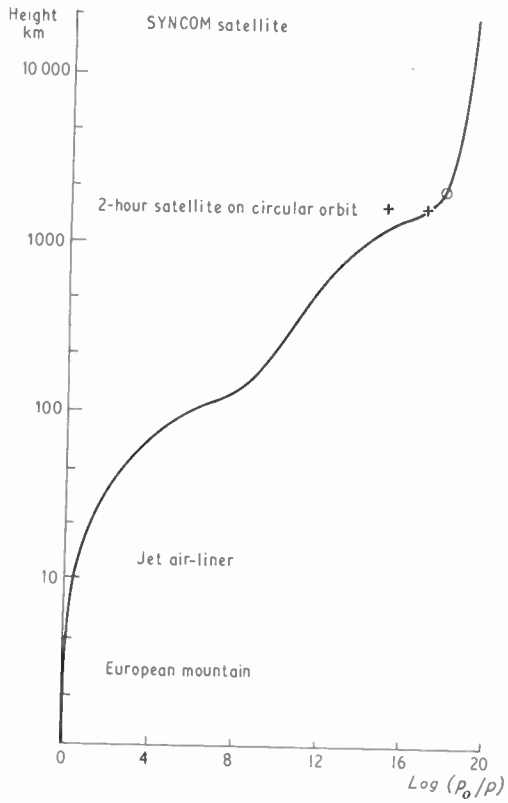

Fig. 1. Atmospheric density versus height. The density is expressed as the logarithm of the ratio of density at a given height to density at sea-level.

to the surrounding plasma this e.m.f. will cause a current to flow; and the energy absorbed by this current must be drawn from the kinetic energy of the moving object. On this basis the atmospheric density must be much less, as indicated by the right-hand cross. The latter is perhaps better since it involves a less sharp curve in order to carry on to the circled point which is taken from Gringauz. It is also supported by the values for the electron density obtained by Bowles[13] from back-scatter radar measurements at 440 MHz, since at this altitude ionization is expected to be fairly complete.

Rocket flights have provided many data in the range 50 to 150 km, but there have been variations of 4 : 1 between individual observations and there has been some controversy as to whether this is due to fluctuations in the structure of the atmosphere or to experimental errors such as gas carried up on the rocket surface; it must be remembered that at 150–200 km we are concerned with pressures less than a millionth of the sea-level atmosphere. King-Hele deduced density in the range of 180–700 km from the slowing down of satellites which entered this part of the atmosphere at perigee (the nearest point on an elliptical orbit), and showed that density variations at 200–300 km consist mainly of a change of the order of $\pm 20\%$ on a 28-day cycle believed to be due to solar disturbances. But at 660 km Martin and Priester[4] (also working from satellite drag) found a density variation of as much as 10 : 1 in a diurnal cycle related to the angle between the satellite's perigee and the Sun's right ascension.

If the Earth's atmosphere were of uniform composition and temperature the pressure would be proportional to density which would be given by $\rho = \rho_0 \exp(-mgh/kT)$ where m is the molecular weight. This is commonly written $\rho = \rho_0 \exp(-h/H)$ where $H = kT/mg$ is the 'scale height'. It can be seen from Fig. 1 that H must vary considerably over the range; and one important reason for this is the dissociation of molecules of O_2 into atomic O at the higher levels. There is also an absence of mixing above a few hundred kilometres which allows the lighter elements to collect at the top. Atomic nitrogen is not commonly found, though the compound NO is important in the E region. Since O is lighter than either N_2 or NO it constitutes the major part of the region from a few hundred to about 1000 km. It is suggested that the change in the curve of Fig. 1 between 1000 and 2000 km is associated with a change of composition from atomic oxygen to atomic hydrogen. As far as can be ascertained,† the temperature remains in the neighbourhood of 200–300°K up to 110 km and this increases to around 1000°K at 300 km and remains practically constant up to 600 km. Some very high electron temperatures have been recorded at still greater heights, but the mean free paths of electrons are there so long that the atoms, or positive ions for the most part, need not be at the same temperature as the electrons.

It is perhaps worth inserting a few scale markers and estimated pressures. A sizeable European mountain is 3 km high, and the air pressure at its summit just over two-thirds of an atmosphere. A jet airliner flies at about 12 km where the pressure is about a fifth of an atmosphere. The Earth's radius is about 6400 km so even at 640 km altitude we are only 1/10th radius away and the effective part of the atmosphere (containing the ionosphere) is quite a thin layer over the Earth. The *Sputnik* type of satellite commonly has a period of the order of two hours, and if its orbit were circular its height would be about 1690 km; but a *Syncom* satellite with a period of 24 hours requires a height of just over 35 800 km. It is only this last type of satellite which is effectively outside the Earth's atmosphere. (See Appendix 1 for calculation of satellite heights from periodicities.)

The pressure in interplanetary space apparently might be of the order of 10^{-13} mm Hg which is a

† See Appendix 2 for sources of data.

good vacuum but not beyond the limit which has been claimed for laboratory vacua, which can extend as far as 10^{-14} mm Hg.

3. Electrical Properties of the Atmosphere

The two atmospheric characteristics which govern radio propagation are the numerical density N_e of electrons (number per cm^3) and the frequency ν with which electrons collide with other particles (molecules, atoms or ions).† The height variations of N_e and ν are sketched in Fig. 2 to logarithmic scales; but the values of N_e are subject to such wide variation both seasonally and with sunspot cycle that the graph is more qualitative than quantitative and it has been broadened into a band in the F region to indicate that the variation there can extend up to 10 : 1. E and F regions can in practice be distinguished, but it is now generally agreed that the fall in N_e between them is of the order of 20% only, which is scarcely perceptible on the scale of Fig. 2.

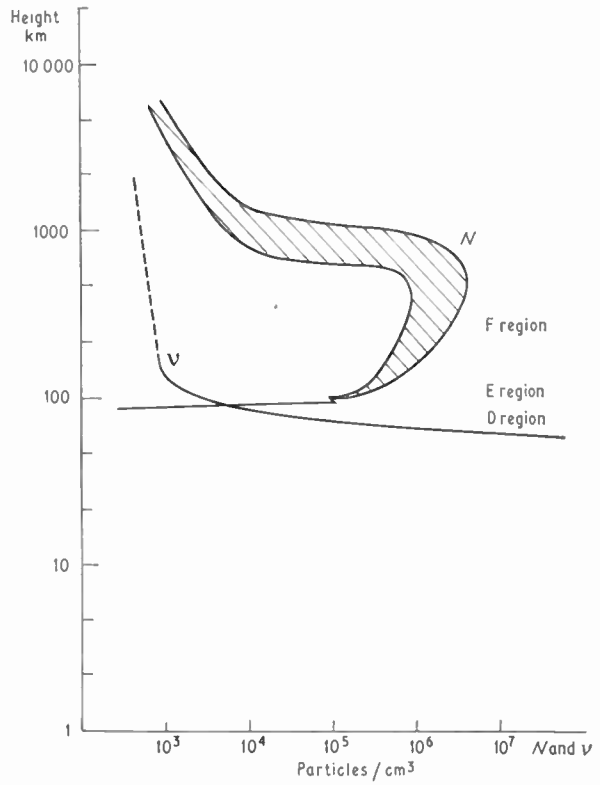

Fig. 2. Numerical density of electrons, and collision frequency of electrons, as a function of height. N and ν are on a logarithmic scale.

† ν is here the total collision of electrons with all other particles; the division between collisions with neutral particles and collisions with ions is discussed later.

These electrical characteristics have been deduced from several radio and electrical measuring techniques as follows:

(i) Vertical-incidence echo-sounding, producing the well-known $h'f$ curves when the radio frequency is varied.[26]

(ii) Radar back-scatter. This utilizes the fact that if electrons move under the influence of the applied field they must give some radiation in the backward direction, and an 'effective scatter cross-section of the electron' can be deduced. (See Section 5.)

(iii) Faraday rotation.[10, 27] A plane-polarized wave is equivalent to the sum of two circularly-polarized waves, one left-handed and one right-handed. In a medium containing electrons subjected to a magnetic field the velocities of left-hand and right-handed waves differ so that they become equivalent to a plane-polarized wave of which the plane of polarization progressively rotates with distance through the medium.

(iv) Ion traps. The ion density in the upper atmosphere can be deduced from the currents flowing to charged probes carried on rockets and satellites.[9]

(v) Capacitance probes.[28] The capacitive reactance of a short aerial depends on the dielectric constant of the medium in which it is placed, and it is shown below that the presence of free electrons in the ionosphere has an effect equivalent to a reduction in the dielectric constant. Hence measurement of the aerial impedance leads to a value for N_e in its vicinity.

The electron density results from a balance between the creation of free electrons by ionization of gas molecules or atoms (mainly by solar radiation) and the loss of electrons either by recombination or by attachment to neutral particles to form negative ions. The collision frequency ν (which is the total frequency with which each electron collides with other particles) can be divided into ν_{en} and ν_{ei} for collisions of electrons with neutral particles and with ions respectively. There is also a coefficient ν_{in} for the collision of ions with neutral particles, but since the movement of ions does not contribute much to high-frequency currents, this coefficient is not important in radio propagation.

At the lower altitudes, where neutral particles are numerous and N_e is relatively small, only ν_{en} is important and is given by

$$\nu_{en} = N_n V_e Q$$

where N_n is the density of neutral particles, V_e the electron velocity and Q the collision cross-section of

the particles for electrons. Provided the electrons are in thermal equilibrium with the gas molecules (which is true if the electrons collide much more frequently with molecules than with each other) V is proportional to $T^{\frac{1}{2}}$; and for a given temperature, N_n is proportional to pressure. It follows that

$$\nu_{en} = \text{constant} \times pT^{\frac{1}{2}} \simeq c'p$$

in the region 50–80 km, where the temperature remains within the range 200–300°K. Huxley[5] proposed the formula $\nu = 1{\cdot}2 \times 10^{-8}p$ with p in mm Hg and this has since been confirmed.[6] In general,[7] on the assumption that all electrons have the mean-square velocity,† the collision frequencies are

$$\nu_{en} = 5{\cdot}4 \times 10^{-10} N_n T^{\frac{1}{2}}$$

$$\nu_{ei} = \{34 + 4{\cdot}18 \log_{10}(T^3/N_e)\} N_e T^{-3/2} \quad \text{.....(1)}$$

In the range from 200 km upwards, the density of neutral atoms is reduced and that of ions increased, so that ν_{ei} becomes comparable with or greater than ν_{en}.

The rate of productions of electron-ion pairs in the atmosphere is proportional to the rate at which radiation is absorbed at that level. This absorption in turn depends on the atmospheric constitution and on the solar spectrum as modified by the atmosphere above the point in question; and in addition the maximum rate of generation is proportional to $\cos \chi$ where χ is the solar zenith angle. This last factor seems to be universally applicable, but the ionization and recombination mechanism vary so much from one region to another that they will be discussed under the different headings.

4. Wave–Electron Interaction

If we ignore for the moment the effect of the Earth's magnetic field, the equation of motion of an electron in an electric field E can be written

$$eE = m\frac{d^2x}{dt^2} + m\frac{dx}{dt} \quad \text{.....(2)}$$

where eE is the applied force, $m\,d^2x/dt^2$ is the acceleration and $m\,dx/dt$ is the rate of loss of momentum through collisions. The equation of a plane wave propagated in the z direction, which in free space would be

$$E = E_0 \cos(\omega t - 2\pi z/\lambda)$$

becomes

$$E = E_0 e^{-kz} \cos(\omega t - \beta z) \quad \text{.....(3)}$$

where

$$\begin{cases} 2k\beta = -\mu N_e^2 \nu\omega/m(\omega^2 + \nu^2) \\ k^2 - \beta^2 = \mu\omega^2[N_e^2/m(\omega^2 + \nu^2) - K_0] \end{cases} \quad \text{.....(4)}$$

k is the attenuation constant, which in the absence of collisions ($\nu = 0$) would be zero. In that case we should have:

$$\beta^2 = \mu_0 k_0 \omega^2 (1 - N_e^2/m\omega^2 K_0) \quad \text{.....(5)}$$

Now the phase-constant β would be ω/c in free space with $c^2 = 1/\mu_0 K_0$ so the phase constant in the medium may be written

$$\beta = (\omega/c)(1 - f_c^2/f^2)^{\frac{1}{2}} \quad \text{.....(6)}$$

where

$f = \omega/2\pi$ and $f_c^2 = N_e^2/(2\pi)^2 m K_0$.

The phase velocity is then increased to the value of

$$V_{ph} = C/(1 - f_c^2/f^2)^{\frac{1}{2}}.$$

Note that this is the phase velocity, and does not conflict with the principle that nothing can be transmitted with velocity greater than c, since the group velocity is in fact reduced to satisfy the requirement that $v_{ph} \times v_g = c^2$. It follows that for f less than $(N_e^2/(2\pi)^2\ m\ K_0)^{\frac{1}{2}}$ the phase constant is imaginary, which means that the wave does not penetrate into such a region.

This can be put in the inverse form that total reflexion at normal incidence will occur when N satisfies the condition

$$N > 4\pi^2 m f^2 K_0/e^2 \quad \text{.....(7)}$$

Inserting numerical values $m = 9{\cdot}1055 \times 10^{-31}$ kg, $K_0 = 1/(36\pi \times 10^9)$ F/m, $e = 1{\cdot}6 \times 10^{-19}$ coulombs, and measuring f in MHz, the required electron density is $N > 1{\cdot}24 \times 10^4 f^2$ electrons per cm^3. For oblique incidence the optical law of total internal reflection gives

$$\mu = \sin i = \cos \varphi \quad \text{.....(8)}$$

where i is the angle of incidence measured from the normal, φ is therefore the angle of elevation above the ground and μ is the refractive index of the medium which is proportional to the square root of the dielectric constant. The critical condition is then

$$N > 1{\cdot}24 \times 10^4 f^2 (1 - \cos \varphi)$$

Consideration of the propagation of a wave in such a medium in terms of Huyghens wavelets suggests that a wave incident at an oblique angle on a medium with N increasing with height will be refracted back again as though it had been reflected by the medium, and the same is in fact true for vertical incidence.

A usual method of exploring the ionosphere is to transmit vertically upwards a series of pulses of gradually increasing frequency. The time taken for a pulse to return is a measure of the apparent height of the plane of reflection, and a plot of this apparent height against frequency is known as an $h'f$ curve. The density N of electrons in general increases with height, so that the pulse will travel some distance through a region where the ionization is appreciable but not yet sufficient to satisfy the condition $f = f_c$ at

† See Section 6 for the effect of velocity distribution.

which the wave is completely reflected. It follows from the previous paragraph that the group velocity will be reduced in this region, and the transit time of the pulse will therefore be greater than it would have been for the same distance in free space, and h' will be greater than the true height at which $f = f_c$. The $h'f$ curve can readily be transformed into a distribution function for the variation of electron density with apparent height, $N(h')$. The conversion to a true $N(h)$ is somewhat complicated, and the easiest approach is to regard the ionosphere as a series of layers within each of which the electron density is constant: methods of converting $N(h')$ to $N(h)$ have been reviewed by Thomas,[8] and the results of some of these were used in Fig. 2.

An alternative method of measuring total electron density along a path is to measure the Faraday rotation of the plane of polarization of a wave. An electron subjected to mutually perpendicular electric and magnetic fields moves in a circle, the periodicity of rotation being governed by the strength of magnetic field. Equating the field acceleration to centrifugal acceleration,

$$eu\mu_0 H_n = mu^2/r \qquad \text{.....(9)}$$

where H_n is the normal component of magnetic field and u the speed of the electron. Since u/r is determined by e and H_n, the number of rotations per second is independent of u. For electrons in the Earth's magnetic field, the number of rotations per second (the gyromagnetic frequency) is about $1{\cdot}3 \times 10^6$ per second, so that there is a resonance effect, with some increase in absorption, when the radio field has a frequency of about 1·3 MHz. A detailed analysis of the propagation of a radio wave in an ionized medium in the presence of a magnetic field leads to the result that the effective dielectric constant of the medium is:

$$K_i = K_0\left[1 - \frac{x}{1 - a \pm (a^2 + y_L^2)^{\frac{1}{2}}}\right] \qquad \text{.....(10)}$$

where

$$x = Ne^2/K_0 m\omega^2 = f_c^2/f^2$$
$$y_L = e\mu_0 H_L/m\omega$$
$$y_T = e\mu_0 H_T/m\omega$$
$$a = y_T^2/2(1-x)$$

H_T and H_L are transverse and longitudinal components of the earth's magnetic field. Equation (10) indicates that the wave is split into two components, that corresponding to taking the positive sign in the denominator of the fraction being called the 'ordinary' wave and that corresponding to the negative sign the 'extraordinary' wave. Due to the circular motion of the electrons in the magnetic field both these waves are elliptically polarized but with opposite directions of rotation.

For high magnetic latitudes y_T is small and the critical frequencies are

$$\left.\begin{aligned} f_c^2 &= f(f+f_H) \text{ ordinary wave} \\ f_c^2 &= f(f-f_H) \text{ extraordinary wave} \end{aligned}\right\}$$

f_H being the gyromagnetic frequency.

Since the refractive indices, and therefore the velocities of propagation, are different for the two components, the major axis of the ellipse, or the plane of polarization in the limiting case that both ellipses are reduced to circles, rotates with distance traversed in proportion to the total electron content of the path. Preferably two frequencies should be transmitted simultaneously to help resolve ambiguities of numbers of complete revolutions. If the transmitter is mounted on a satellite a record of rotation against height will give a profile of the electron density. Since Faraday rotation is not reversed in direction when the direction of transmission is reversed, echo techniques can also be used: the rotation of polarization of Moon echoes received at Jodrell Bank has been used to measure the total electron content of the atmosphere. The rotation is a function both of N and ω, so that by transmitting simultaneously two different frequencies between a satellite and ground it is possible to eliminate ω and determine the integrated value of N over the path covered. The distribution of electrons versus height can be found by difference of observations at different heights. The results obtained by this method (e.g. those reported by Garriott[10]) are in good agreement with those deduced by other methods.

The theoretical formula for the attenuation in the presence of a magnetic field is [11]

$$k_m = \frac{k_0}{f^2}\left[p + \frac{2p(p-1)^2}{(p^2-1) - p^2\tau^2}\right] \qquad \text{.....(11)}$$

where k_0 is the absorption that would be found at 1 MHz in the absence of a magnetic field, f is measured in MHz, τ is the ratio of the gyromagnetic frequency to the working frequency, θ is the angle between the direction of propagation and the magnetic field and p is a solution of

$$p^2(1-\tau^2) - p(2-\tau^2\sin^2\theta) + 1 = 0 \qquad \text{.....(12)}$$

Being a quadratic in p, equation (12) must have two roots which will apply to the ordinary and extraordinary wave, the absorption being greater for the extraordinary wave.

Note that in general the attenuation indicated by equation (11) varies inversely as the square of the frequency, so that for propagation via the ionosphere one should use the highest frequency consistent with obtaining the desired reflexion.

5. Incoherent Scatter from Electrons

In determining the phase constant or apparent permittivity of the ionosphere one writes $J_0 = Neu$ where u is the drift component of velocity common to all the electrons and one ignores the actual behaviour of individual electrons. But if one looks on a finer scale, each electron which is accelerated must in consequence radiate the same kind of pattern as a doublet, i.e. with field-strength proportional to sin θ where θ is the angle between direction of observation of the field and the direction of acceleration. Since the field which accelerates the electron is transverse to the direction of propagation, the radiation from the electron will be uniform in the plane of incidence, and some will be scattered back. (The part which is scattered forward will mingle with the main wave propagation.)

The radiation from an accelerated electron is calculated from classical electromagnetic theory along the following lines. The surrounding space is occupied by flux of electric displacement $D = e/4\pi r^2$. If the electron is in steady motion there will be throughout the space a displacement current of density $J = \mathrm{d}D/\mathrm{d}t$. Corresponding to this current there will be at every point, including the position of the electron, a magnetic vector potential calculable from the formula

$$A = \int_{\bullet} \frac{J\mathrm{d}V}{r}$$

where the volume integral is to be taken over all space. A is proportional to $\mathrm{d}D/\mathrm{d}t$ and hence to velocity, so if further the velocity changes there will be a corresponding value of $\mathrm{d}A/\mathrm{d}t$. But $\mathrm{d}A/\mathrm{d}t$ is equivalent to an electrical field E which will exert a force on the charged electron. Hence work has to be done in accelerating the electron (apart from accelerating any possible 'mechanical' mass) and this work must be transformed into radiation. The intensity of the radiation is calculated to be

$$I(\theta) = (1/\mu)\frac{e\ddot{x}\sin\theta}{4\pi\varepsilon_0 c^2} \text{ watt/steradian}$$

where θ is the angle between the direction of observation of the radiation and the direction of acceleration $\ddot{x}$, and μ is the intrinsic impedance of free space. Since $\ddot{x} = eE/m$ it can be eliminated in terms of the incident energy flux $\Phi_0 = E^2/\mu$ watts/m^2 and then

$$I(\theta) = \Phi_0\left(\frac{e^2}{4\pi m\varepsilon_0 c^2}\right)^2 \sin^2\theta = \Phi_0 r_e^2 \sin^2\theta$$

where $r_e = 2{\cdot}82\times10^{-15}$ metre is the 'classical radius of the electron' and is such that if the charge on the electron is contained within this radius, then the electromagnetic momentum accounts for the observed mass of the electron. Proceeding a stage further, if one compares the total re-radiated energy with the incident energy flux, one can deduce an equivalent scattering cross-section of

$$\sigma_e = 6{\cdot}65\times10^{-29}\text{m}^2.$$

Since the electrons scatter individually, their scattering is said to be incoherent and the total scattered intensity is proportional to the number of electrons. Therefore, if one directs a suitable radar beam vertically upwards, there should be reflections from all heights above 60 km, the intensity at any height being proportional to the local electron density. The frequency used is too high to be reflected coherently by the ionosphere, so that this method is able to observe the decreasing electron density in the valley between E and F, and again beyond F2, regions which otherwise are accessible only to methods such as Faraday rotations which measure the integrated electron density over a complete trajectory. A typical radar equipment for this purpose operates at 41 MHz with 100 μs pulses of peak power of one megawatt and an antenna cross-section of $1{\cdot}6\times10^4$m^2.

6. The D Region

Figure 2 suggests that there is little ionization below 85 km but more detailed examination shows that during the day there are electron densities of the order of 200 to 1000 per cm^3 in the range from 65 to 100 km. This density is not sufficient to reflect frequencies in the megahertz range, and since the collision frequency of the electrons is between 10^6 and 10^7 per second, the attenuation of such frequencies is high, accounting for the failure of medium-frequency waves to propagate via the ionosphere in day-time. On the other hand very low frequencies, say 15 kHz, are reflected without appreciable attenuation. Since the wavelengths are of the same order as the height of reflection (15 kHz corresponds to 20 km wavelength) the propagation is often thought of as using waveguide modes between the earth and the ionosphere. This v.l.f. band is the only form of long-range radiocommunication which is fully reliable in an absolute, rather than statistical, sense. (Its disadvantage is that aerial radiation efficiencies are only a few per cent so that powers up to 1000 kW c.w. are used, and of course the number of separate channels which can be made available is small.) For moderate distances (400–900 km) equivalent reflection coefficients have been calculated theoretically[14] to be generally between 1·0 and 0·5 for the frequency range 100 to 10^5 Hz, but experimental results[15] are rather less favourable as shown in Table 1.

In the classical Appleton–Hartree–Booker theory of magneto-ionic propagation it was implied that all electrons have the same collision frequency, i.e.

that they all have the same thermal-agitation velocity, whereas in fact they will have a Maxwell-Boltzmann distribution at the higher gas pressures and may have a non-equilibrium distribution at very low pressures. Since v enters the phase-velocity formula through a term $v^2-\omega^2$ its exact value will not be critical when ω is 1 MHz or more and v is of the order of 10^3–10^4.

Table 1

D region propagation characteristics

Season and time	Frequency (approximate) 16 kHz		70 kHz	
	R	h, km	R	h, km
Summer				
noon	0·1	74	0·002	70
night	0·5	90		
Winter				
noon	0·25	79	0·03	90
night	0·35	92	0·33	90
Day s.i.d.	0·09	$\Delta h \simeq 6$	0·001	$\Delta h \simeq 6$

R = equivalent reflexion coefficient

h = apparent reflexion height

Δh = change in h on occurrence of s.i.d.

s.i.d. = sudden ionospheric disturbance (solar flare)

But in v.l.f. transmission the situation is reversed, since $\omega \simeq 10^5$ and v is 10^6–10^7, so the correction for the velocity distribution function can be significant. On this basis Kane[16] finds that the data obtained by Seddon[17] from a rocket flight at White Sands, New Mexico, at noon on 29th June 1956 indicate an electron density showing a reasonably steady trend approximately from $10^3/cm^3$ at 56 km, to $6\times10^3/cm^3$ at 70 km

The day-time D layer is predominantly due to solar radiation.[18] Molecular O_2 and N_2 are ionized by Lyman α radiation (1216 Å) and to a less extent by cosmic rays. Above 85 km there is some contribution also from Lyman β (1026 Å) and from x-rays between 30 and 40 Å, but this height is approaching the E region. Since there is no solar radiation at night, the lower part of the night E region must presumably serve as reflector for the frequencies usually regarded as propagating via the D region.

The intense radiation from solar flares can cause the ionization to extend down below 50 km and lead to severe absorption of frequencies from 100 kHz up; but as shown in Table 1, it has little effect on v.l.f. transmissions at 15–30 kHz.

7. The E Region

The E region is at a height of 80–110 km, and the day-time value of electron density is of the order of 10^5 per cm^3. According to the simplest (Chapman) theory,[7] the rate of production of ions and electrons is proportional to $\cos\chi$, where χ is the solar zenith distance, and the rate of loss by recombination is αN^2 where α is the coefficient of recombination. The net rate of change of electron density is then

$$\frac{dN}{dt} = q_0 \cos\chi - \alpha N^2$$

with q_0 the maximum rate of generation. In a completely steady state one could write

$$N = \sqrt{\{(q_0 \cos\chi)/\alpha\}}$$

Since the critical radio frequency f_E for reflection from the E region is proportional to $\sqrt{N}$, this would lead to

$$f_E^4/\cos\chi = \text{constant}$$

This can be put in the alternative form

$$f_E = K(\cos\chi)^p$$

where p would ideally have the value $\frac{1}{4}$ but experimentally has values ranging from about 0·22 to 0·3 according to latitude and sunspot cycle. The latitude effect is the more pronounced, p having a larger value at the equator than at the poles. The effect of sunspot number R can be represented by

$$f_E^2 = a_\mu(1+0{\cdot}004R)$$

Since R is normally within the range 1 to 200, this represents a variation of f_E of less than $\pm 20\%$.

The ionizing agent for the day-time E region consists of soft x-rays, and the best estimate of the recombination coefficient is the range $1{\cdot}5\times10^{-8}$ to 4×10^{-8} $cm^{-3}\ s^{-1}$. (See Appleton's review paper.[19])

Since the exponential decay of the layer after removal of the source of ionization (i.e. after sunset on the above theory) can be written

$$\frac{1}{N}\cdot\frac{dN}{dt} = -\alpha N$$

and α and N are of order 10^{-8} and 10^5 respectively, the E region should fade away in an hour or two and some other explanation must be found for the persistence of a fairly steady level of ionization throughout the night. The source of this night-time E and of sporadic E which occur in various forms throughout the 24 hours is not entirely clear. It is, however, certain that meteorites make a substantial contribution, since 100 km is about the height at which the atmosphere is dense enough to cause them to burn up and leave a trail of ionized gas. Manning and Eshleman[20] reported that meteors of 5th magnitude, which is barely visible, occur on the average at one every 60 seconds in a sky area of 100 km by 100 km; but 15th magnitude meteors, the smallest detectable with their 30 MHz radar at a distance of 150 km, occur

at average intervals of 0·006 second. As the duration for small meteor trails is about 0·5 second, this gives continuous coverage if one averages over a fairly large area. In addition there is the rather baffling phenomenon known as 'sporadic E' ionization: baffling because it occurs in various 'forms' described by the associated *h'f* curves and there is no obvious explanation for any of these, but one form or another occurs for about 50% of the time in most parts of the world. For an electron density of 10^5 the vertical-incidence critical frequency should be under 3 MHz, but reflections up to 5 MHz have been found and 50 MHz television signals have been reflected at grazing incidence to propagate thousands of miles. (This must not be confused with propagation of television signals over 100 miles or so by tropospheric refraction.)

It appears that sporadic E reflection is often caused by thin patches of relatively intense ionization which are sharply defined in height; and that some of them are correlated with wind shear. 'Wind shear' is the change with height of the direction and speed of the wind in the ionosphere, the typical speeds being tens of metres per second in the neighbourhood of 100 km; and it has been verified from optical observation of meteor trails and artificial flares that this does represent a bodily movement of gas, and not the propagation of a fluctuation in the ionization density of a stationary atmosphere. The general wind drifts have been determined by means of radar reflection from meteor trails[21] and show a distinct 12-hour cycle which may be attributed jointly to tidal and solar-heating effects. More detailed examination[22] indicates that wind shear may be as high as 50 $m.s^{-1}$ km^{-1} and that the effect is correlated over a distance of about 160 km. Although sporadic E correlates in height with the maximum wind shear, there is evidence that at least the 'equatorial' type arises from irregularities which are 100–300 m long in the direction of the Earth's magnetic field but of the order of 1 m only in width.

There is a theory[23] that irregularities in the ionosphere are caused by atmospheric waves of a hydrodynamic, rather than acoustic, type which travel upwards: since they tend to carry constant energy the amplitude increases with height as the pressure falls. The dominant period is calculated to be 200 minutes, and the smallest disturbance that could continue above 100 km is estimated to be 1 km wide. Unfortunately this does not accord either with the small size which many sporadic E patches appear to have or with the highest values of wind shear.

There appears at present to be no satisfactory explanation for the various forms of sporadic E which are found, but they occur sufficiently frequently to be a significant factor in radio communications. The possible mechanisms of ionospheric scatter propagation have been reviewed by Bowles.[13]

8. The F Region

Although the electron density varies fairly smoothly throughout the F region from some 130 km to the outside of the atmosphere, there are valid reasons for distinguishing F1 and F2 in terms of their physical mechanisms and patterns of variation. The brief decrease in N_e between E and F1 was difficult to evaluate by radio echo sounding, but in the light of additional information obtained from rockets and satellites it appears that the minimum here is only about 20% below the E maximum.

The F1 layer appears to lose electrons by recombination so that $dN/dt = q - \alpha N^2$, and has a relaxation time of about 4 minutes. A peak density quoted by Martyn[24] for noon at the equator at sunspot minimum is $2{\cdot}5 \times 10^5$ electrons per cm^3 and the sunspot effect is quite small, according to the formula

$$N_{mF1} = 2{\cdot}5 \times 10^5(1 + 0{\cdot}0062\,R)$$

where N_{mF1} is the peak number of electrons/cm^3 in F1. The intensity is naturally quite small after dark, falling to about 10% of the daylight value, so that F2 is then dominant when both E and F1 have sunk to low values. At the height of 150 to 200 km the ions found in rocket sampling experiments were divided in comparable proportions between molecular oxygen (O_2^+), nitric oxide (NO^+) and atomic oxygen (O^+), but above 250 km atomic oxygen predominates.

The F2 region shows greater variation with the sunspot cycle than other regions and the variation of maximum ionization density at noon is represented approximately by

$$N = N_0(1 + 0{\cdot}02\,R)$$

where N_0 is the value at sunspot minimum. Remembering that R can be as much as 200, this implies a five-fold variation in relation to sunspot only. The fact that the maximum electron density of F2 varies with sunspot cycle so much more than that of other regions, and that the height at which it occurs also shifts between about 250 km at sunspot minimum and 350 km at sunspot maximum, suggests that the radiation involved differs from that involved at lower levels. A distinctive feature of the F2 region is that the latitudinal distribution of equilibrium electron density varies more nearly as $\cos\chi$ than as $(\cos\chi)^{\frac{1}{2}}$. The rate of loss of electrons proportional to N corresponds to an attachment-like mechanism in which electrons are lost by interaction with ions which are present in large numbers, independently of the number of electrons. (Such positive ions could arise without the liberation of electrons by the dissociation of a molecule into two parts, with positive and negative charges respectively.)

9. Scatter Propagation

Where there is an inhomogeneity in the transmission medium there must be some degree of scattering, whether or not the inhomogeneity is sufficient in intensity and extent to cause specular reflection. Inhomogeneities in the lower atmosphere cause 'tropospheric scatter' of frequencies which are not propagated via the ionosphere, and they thus assist transmission beyond the v.h.f. horizon. (This differs from the optical horizon because the normal temperature and pressure gradients in the atmosphere cause the dielectric constant to vary with height; the result is equivalent to increasing the Earth's radius by a factor of 4/3 and extending the horizon distance correspondingly.)

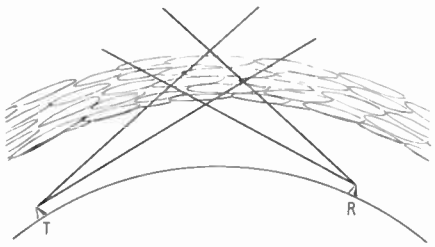

Fig. 3. Schematic diagram of propagation path in an ionospheric scatter link.

In ionospheric scatter communication the technique is to use high-gain antennas at both transmitter and receiver and direct their beams to a common region in the ionosphere so that any scattering in that region will deflect some energy from the transmitting beam to the receiving beam (Fig. 3). The operating frequency is naturally higher than would be specularly reflected from the ionosphere, and Bowles[13] suggests the following specification for a v.h.f. scatter link utilizing scattering from the D region:

Antenna gain, referred to isotropic	25dB
Transmitter power	40 kW
Receiver bandwidth	6 kHz
Path length	1000–2300 km

The scattering follows closely a power law of the working frequency:

$$P_R/P_T \propto f^{-m}$$

where m varies a little with conditions but averages 7·5.

The ionized trails from meteors make a considerable contribution to scatter propagation. But there is also scatter from irregularities in the D region, which are attributed to turbulence, and this component is strongest at noon when the solar ionization of the region is a maximum. (For details see Bowles.[13])

Forward scattering is also obtained from irregularities in the F region. These irregularities can be observed on *h'f* records obtained by the usual vertical echo-sounding technique, and are also responsible for the scintillation of radio stars.

10. Acknowledgment

This paper is based on part of a lecture course given while the author was Visiting Professor of Telecommunications at McGill University.

11. References

1. S. Chapman, 'The Earth and its environment', *Proc. Inst. Radio Engrs*, **47**, p. 137, February 1959.
2. N. Nicolet, 'The constitution and composition of the upper atmosphere', *ibid.*, p. 142.
3. D. G. King-Hele, 'Density of the upper atmosphere from analysis of satellite orbits: further results', *Nature*, **184**, No. 4695, p. 1267, 24th October 1959.
4. H. A. Martin and W. Priester, 'Measurement of solar and diurnal effects in the high atmosphere by artificial satellites', *Nature*, **185**, No. 4713, p. 660, 27th February 1960.
5. L. G. H. Huxley, Communication to XIIth General Assembly of the U.R.S.I., 1957.
6. J. C. Seddon, 'Summary of rocket and satellite observations related to the ionosphere', Proceedings of U.R.S.I., 13th General Assembly, London, 1960.
7. S. Chapman, 'The electrical conductivity of the ionosphere: A review', *Nuovo Cimento*, **4**, Series 10, Supp. No. 4, p. 1385, 1956.
8. J. O. Thomas, 'The distribution of electrons in the ionosphere', *Proc. I.R.E.*, **47**, p. 162, February 1959.
9. K. I. Gringauz, 'Determining local concentrations o charged particles in the ionosphere and interplanetary space', *J. Brit. Instn Radio Engrs*, **22**, No. 5, p. 419, November 1961.
10. O. K. Garriott, 'The determination of ionospheric electron content and distribution from satellite observations. I. Theory of analysis, II. Results of analysis, *J. Geophys. Res.*, **65**, pp. 1139 and 1151, April 1960.
11. 'Ionospheric Radio Propagation', N.B.S. Circular 462, U.S. Government Printing Office, 1948.
12. S. D. Drell, H. M. Foley and M. A. Ruderman, 'Drag and propulsion of large satellites in the ionosphere: an Alfven propulsion engine in space', *Phys. Rev. Letters*, **14**, No. 6, p. 171, 8th February 1965.
13. K. L. Bowles, 'Radio wave scattering in the ionosphere', *Advances in Electronics and Electron Physics*, **19**, p. 55, 1964.
14. J. R. Johler and Lillie C. Walters, 'On the theory of reflexion of low and very-low-radiofrequency waves from the ionosphere', *J. Res. Nat. Bur. Stands.*, **64D**, p. 269, May–June 1960.
15. R. N. Bracewell, K. G. Budden, J. A. Ratcliffe, T. W. Straker and K. Weekes, 'The ionospheric propagation of low-and-very-low frequency waves over distances less than 1000 km', *J. Instn Elect. Engrs*, **98**, Pt. III, p. 221, May 1951.
16. J. A. Kane, 'Re-evaluation of ionospheric electron densities and collision frequencies derived from rocket measurements of refractive index and attenuation', *J. Atmos. Terrestr. Phys*, **23**, p. 338, December 1961.

17. A. J. C. Seddon, 'Differential absorption in the D and lower E regions', *J. Geophys. Res.*, **63**, p. 209, March 1958.
18. A. C. Aikin, 'The sunrise absorption effect observed at low frequencies', *J. Atmos. Terrestr. Phys*, **23**, p. 287, December 1962.
19. E. V. Appleton, 'The normal E region of the ionosphere', *Proc. I.R.E.*, **47**, p. 155, February 1959.
20. L. A. Manning and V. R. Eshleman, 'Meteors in the ionosphere', *Proc. I.R.E.*, **47**, p. 186, February 1959.
21. J. S. Greenhow and E. L. Neufield, 'The height variation of upper atmospheric winds', *Phil. Mag.*, **1**, 8th series, p. 1157, December 1956.
22. R. H. Briggs, 'Survey of ionospheric drifts', in 'Ionospheric Radio'. Ed. W. G. J. Beynon (Elsevier, Amsterdam, 1962).
23. C. O. Hines, 'An interpretation of certain ionospheric motions in terms of atmospheric waves', *J. Geophys. Res.*, **64**, p. 2210, December 1959.
24. D. F. Martyn, 'The normal F region of the ionosphere' *Proc. I.R.E.*, **47**, p. 146, February 1959.
25. T. E. Van Zandt and K. L. Bowles, 'Use of the incoherent scatter technique to obtain ionospheric temperature', *J. Geophys. Res.*, **65**, p. 2627, September 1960.
26. J. N. Brown, 'Automatic sweep-frequency ionosphere recorder, model C-4', *Proc. I.R.E.*, **47**, p. 296, February 1959.
27. Warren W. Berning, 'Earth satellite observations of the ionosphere', *Proc. I.R.E.*, **47**, p. 280, February 1959.
28. J. E. Jackson and J. A. Kane, 'Performance of an r.f. impedance probe in the ionosphere', *J. Geophys. Res.*, **65**, p. 2209, July 1960.
29. C. D. Hodgman (Ed.) 'Handbook of Chemistry and Physics', 46th edition (Chemical Rubber Publishing Co., New York, 1965).

12. Appendix 1

Height of satellite of given period

If the orbit is circular its radius is fixed by equating gravitational to centrifugal force:

$$mr\omega^2 = mg$$

where m = mass of object, r = radius of orbit in metres, ω = angular velocity in radians per second and g = gravitational acceleration in metres/second^2 at radius r. To the degree of approximation that the Earth is regarded as a sphere:

$$g = (r_0/r)^2 g_0$$

where r_0 = radius of the Earth's surface and g_0 is the value of g at r_0. It follows from the above two equations that

$$r = [g_0(r_0/\omega)^2]^{\frac{1}{3}}$$

Substitution of $g_0 = 9{\cdot}81$ m/s^2 and $r_0 = 6{\cdot}37 \times 10^6$m, leads to the results given in Section 2.

13. Appendix 2

The temperature of the ionosphere

Martyn[24] stated that it was well established that the temperature is about 300°K at 110 km, and from variation of particle density with height deduced a temperature of 560°K at 160 km. The 'Handbook of Chemistry and Physics'[29] gives a graph (page F.114) showing kinetic temperature rising from about 670°K at 175 km to 950°K at 300 km. Van Zandt and Bowles[25] deduced from electron-backscatter that the atmosphere between 370 and 520 km is isothermal with temperature $1050° \pm 90°$K. They claimed that this agreed with the value deduced from the line width of 6300 Å nightglow.

Manuscript first received by the Institution on 26th April 1965 and in final form on 10th January 1966 (Paper No. 1070.)

U.D.C. 621.372.5

Pulse Response of Delay Lines

II. *m*-derived Delay Lines

By

M. AVINOR, Ph.D., C.Eng.
(*Associate Member*)†

AND

Y. PARTOM, M.Sc.†

Summary: Numerical solutions have been obtained for unit step response of *m*-derived delay lines in which inductive coupling exists between adjacent sections. It is shown that for optimum coupling both rise-time and overshoot are improved in comparison to constant-*k* delay lines.

1. Introduction

Calculation of the response of *m*-derived delay lines may be approached along the same lines developed in the previous paper.[1] However, the case of the infinite line proved to be more difficult to deal with. The terminated finite line, on the other hand, yields easily to the same general program.

2. Line of Infinite Length

Figure 1 shows one section of an infinite line with mutual inductive coupling between the inductances. Let the total inductance, measured across the upper terminals, be $2L$. The mutual inductive coupling coefficient k may vary from -1 to $+1$, depending on whether the coils are series opposing or series aiding. The most useful delay responses are obtained for positive values of k, and only these cases will be considered in this paper.

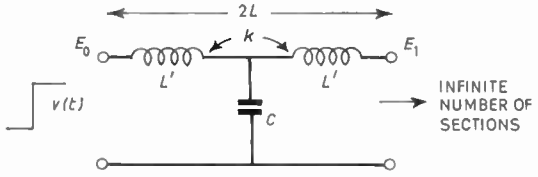

Fig. 1. First section of an *m*-derived delay line with lumped constants.

The coupling coefficient may be obtained experimentally by measuring the inductance L' of each individual coil and the total inductance $2L$. Then:

$$k = L/L' - 1 \qquad \text{.....(1)}$$

and

$$m = \frac{1+k}{1-k} \qquad \text{.....(2)}$$

For the calculation of the Laplace transform it is convenient to transform the network of Fig. 1 to the equivalent network of Fig. 2.

† Scientific Department, Ministry of Defence, Israel.

For delay lines the inductance l in Fig. 2 is a negative quantity, and cannot be realized physically as a distinct component.

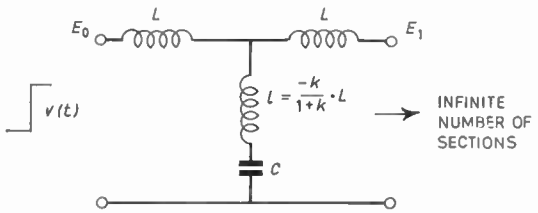

Fig. 2. Equivalent circuit of the *m*-derived delay line section of Fig. 1.

Writing the transmission matrix of the network of Fig. 2 and using the fact that the output impedance is equal to the input impedance, the transform of the line impedance is obtained as:

$$\mathscr{L}(z) = L\sqrt{\alpha_K^2+s^2} \cdot \frac{1}{\sqrt{m}} = \sqrt{L(2l+L)} \cdot \sqrt{\alpha_K^2+s^2} \qquad (3)$$

where

$$\alpha_K^2 = \frac{2m}{LC} = \frac{2}{(L+2l)C}$$

After the transients die out the line impedance becomes $(2L/C)^{\frac{1}{2}}$, as usual. Using eqn. (3) in the solution of the network, the Laplace transform of the output voltage is obtained:

$$\mathscr{L}\left(\frac{E_1}{E_0}\right) = \frac{1}{s} \cdot \frac{(\sqrt{\alpha_K^2+s^2} - s\sqrt{m})^2}{\alpha_K^2(1+s^2 . l . c)} \qquad \text{.....(4)}$$

where E_0 is the magnitude of the input step.

The transform of the voltage tapped off after n sections is similarly

$$\mathscr{L}\left(\frac{E_n}{E_0}\right) = \frac{1}{s} \cdot \frac{(\sqrt{\alpha_K^2+s^2} - s\sqrt{m})^{2n}}{\alpha_K^{2n}(1+s^2(1-m)/\alpha_K^2)^n}$$

$$= \frac{1}{s}\left[\frac{\sqrt{\alpha_K^2+s^2} - s\sqrt{m}}{\sqrt{\alpha_K^2+s^2} + s\sqrt{m}}\right]^n \qquad \text{.....(5)}$$

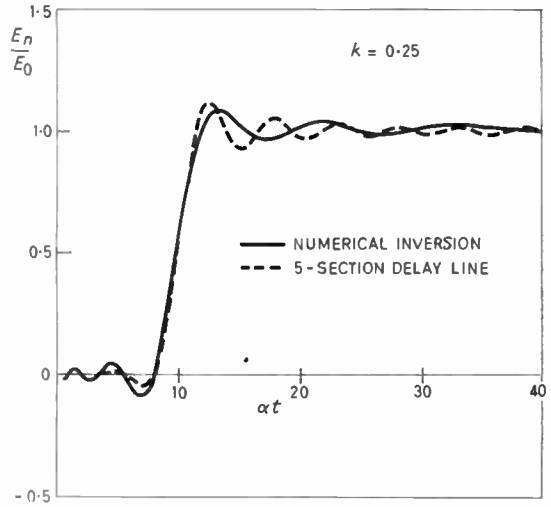

Fig. 3. Voltage output after five sections of an infinite delay line (full line) obtained by numerical inversion of the Laplace transform. Compared to the output of a five-section finite line terminated at both ends (dashed line).

On putting $m = 1$ (i.e. $k = 0$, no coupling) in eqns. (3), (4) and (5) they reduce to the corresponding equations of the constant-k delay lines. Inversion of eqn. (5) in closed form seems to be impossible. Therefore a direct numerical inversion of the Laplace transform was attempted.[2] The main drawback of numerical inversion methods is the high numerical accuracy required for even moderate precision in the results. A computer program has been prepared for the numerical inversion of eqn. (5). All the critical calculations were carried out in double precision arithmetic (i.e. about 22 decimal significant figures) and the final series contained 33 terms. The results were checked by putting $k = 0$ and comparing with the Bessel function summation. The numerical inversion agreed with the exact calculation within about 10%. The results thus give only a rough estimate of the function and the response obtained for a five-section delay line is shown in Fig. 3.

In an infinite delay line the coupling between the sections introduces an auto-transformer action along the line. Therefore some output appears immediately on applying the input pulse. The magnitude of the initial amplitude may be obtained as usual by dropping the $1/s$ term and letting s tend to infinity.

The result is:

$$\lim_{t\to 0}\frac{E_n}{E_0}=\left[\frac{\sqrt{1-k}-\sqrt{1+k}}{\sqrt{1-k}+\sqrt{1+k}}\right]^n=\left[\frac{1-\sqrt{m}}{1+\sqrt{m}}\right]^n \quad \ldots\ldots(6)$$

Equation (6) shows that the initial output may have either the polarity of the input pulse or the inverse polarity, depending on whether the number of sections is even or odd. As the number of sections increases this initial output quickly decreases.

3. Line of Finite Length

Evaluation of the transform polynomials of a finite line, and their inversion by summation of the residues proceeds exactly as for the constant-k delay line. The main difference is that the coupling adds an additional term in the numerator of the Laplace transform. The numerical calculation, however, remains essentially the same. Let one section of the delay line be as in Fig. 4, and its equivalent network as in Fig. 5.

The Laplace transform of the output voltage is:

$$\mathscr{L}\left(\frac{E_1}{E_0}\right)=\frac{R}{s}\cdot\frac{sl+\dfrac{1}{Cs}}{PA_2}$$

$$=\frac{R}{Ls}\cdot\frac{s^2\dfrac{l}{L}+\dfrac{R^2}{2L^2}}{s^3\dfrac{L+2l}{L}+s^2\dfrac{R}{L}\cdot\dfrac{L+l}{L}+s\dfrac{R^2}{L^2}+\dfrac{R^3}{2L^3}} \quad \ldots\ldots(7)$$

where

$$PA_2=\begin{vmatrix} s(L+l)+\dfrac{1}{Cs} & -sl-\dfrac{1}{Cs} \\ -sl-\dfrac{1}{Cs} & s(L+l)+\dfrac{1}{Cs}+R \end{vmatrix}$$

For n sections the general transform is:

$$\mathscr{L}\left(\frac{E_n}{E_0}\right)=\frac{R}{s}\cdot\frac{\left(sl+\dfrac{1}{Cs}\right)^n}{PA_{n+1}} \quad \ldots\ldots(8)$$

The determinants PA_n are seen to differ from the corresponding ones for the constant-k delay lines in having the term $(L + l)$ instead of L on the main

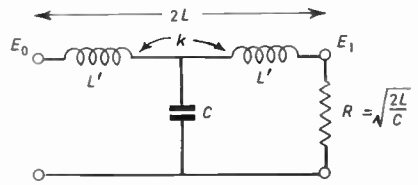

Fig. 4. One section of m-derived delay line terminated with a resistance equal to the characteristic impedance of the line.

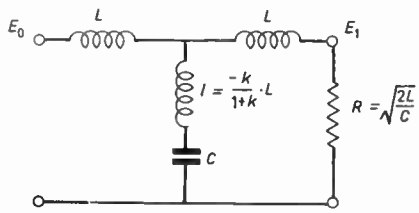

Fig. 5. Equivalent circuit of one-section m-derived delay line of Fig. 4.

diagonal, and $-sl-1/Cs$ on the adjacent diagonals instead of $-1/Cs$. The same recursion formulae developed in the previous paper hold here too.

Inserting convenient numbers for R, L and C (2, 2 and 1 respectively) the Laplace transform of the output voltage takes the final form:

$$\mathscr{L}\left(\frac{E_n}{E_0}\right) = \frac{(ls^2+1)^n}{s(\text{polynomial of } s)} \qquad \ldots\ldots(9)$$

Calculation of the coefficients of the denominator polynomials was carried out exactly as before, using the same recursion formulas. Factorizing the polynomials and inversion by summation of the residues is then quite straightforward.

The coefficients of the denominator polynomials in s, for various degrees of coupling, are given in Table 1 for a five-section delay line. When the coupling is not zero the coefficients are no longer all integers. After inversion of the Laplace transform the results are obtained in the form of a finite trigonometric series:

$$E_n/E_0 = 1 - K.\exp(X.\alpha t) + \sum_{j=1}^{n} \exp Y_j.\alpha t(L_j \cos W_j.\alpha t - M_j \sin W_j.\alpha t) \qquad \ldots\ldots(10)$$

where

$$\alpha = (2/LC)^{\frac{1}{2}} = R/L.$$

The constants for a five-section delay line, for various degrees of coupling, are given in Table 2.

Table 1

Coefficients of the network polynomials in s for a five-section delay line terminated at both ends (eqn. 9)

Power of s	Coupling coefficient k			
	0	0·2	0·3	0·4
0	2	2	2	2
1	20	20	20	20
2	100	96·66	95·38	94·28
3	330	300	288·46	278·57
4	800	668·89	619·64	577·91
5	1504	1135·11	1002·70	893·39
6	2240	1505·92	1258·15	1060·76
7	2720	1601·48	1238·45	989·39
8	2560	1318·64	964·67	710·10
9	2048	897·63	604·82	408·68
10	1024	391·15	245·46	153·84
11	512	159·69	90·06	50·36

Table 2

Values of the constants in eqn. (10) for a five-section delay line terminated at both ends. Normalized to give unity final output

k	K	L	M	X	Y	W
0	−3·16328	−0·74375	−0·21932	−0·35761	−0·15370	0·64826
		0·76849	4·82496		−0·32816	0·21470
		0·16328	0·14357		−0·07119	0·83307
		0·20000	−1·15470		−0·25000	0·43301
		−0·02474	−0·03256		−0·01814	0·95671
0·2	−3·85088	−1·12979	−0·22298	−0·37806	−0·21765	0·70163
		2·56115	−1·73607		−0·30065	0·44401
		0·28973	0·20535		−0·12268	0·95665
		1·17943	5·97357		−0·35788	0·21190
		−0·04964	−0·05119		−0·03684	1·15125
0·3	−4·29039	−1·39306	−0·20887	−0·38912	−0·25555	0·72459
		2·90428	−2·13100		−0·32692	0·44711
		0·38613	0·24333		−0·15978	1·02161
		1·46376	6·69838		−0·37310	0·21025
		−0·07070	−0·06350		−0·05285	1·26649
0·4	−4·83096	−1·73004	−0·17571	−0·40113	−0·29838	0·74456
		3·31167	−2·63725		−0·35439	0·44865
		0·51929	0·28742		−0·20820	1·08900
		1·83209	7·58242		−0·38908	0·20842
		−0·10205	−0·07823		−0·07691	1·39996

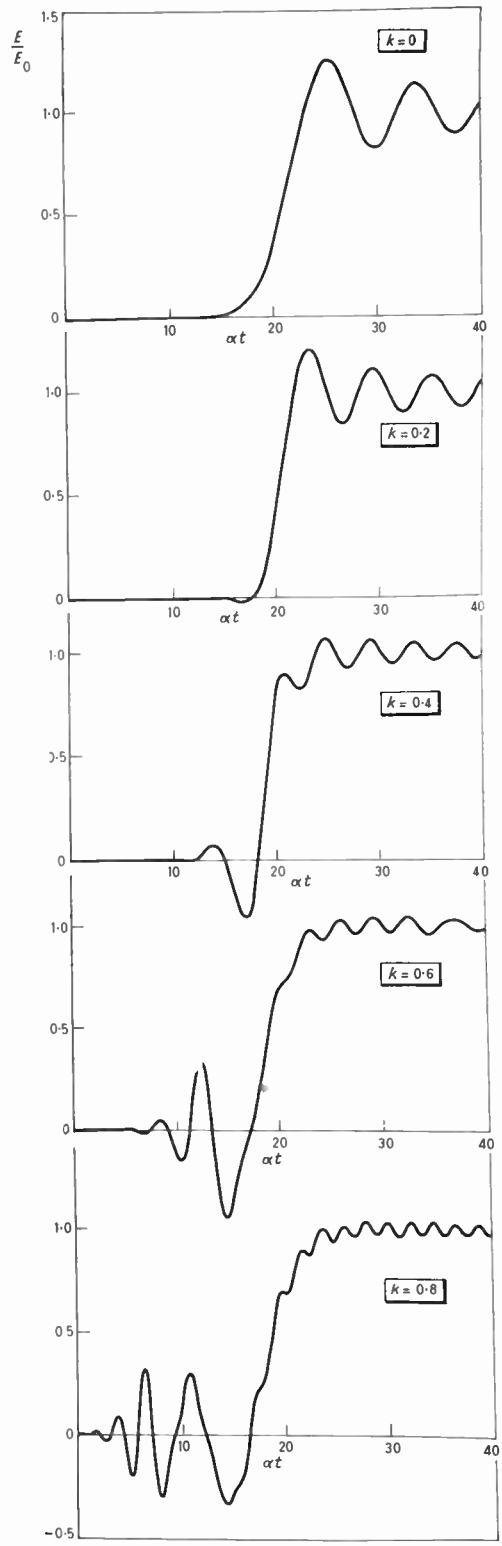

Fig. 6. Response of a ten-section *m*-derived delay line to unit step excitation with various degrees of coupling. Step generator of zero internal impedance.

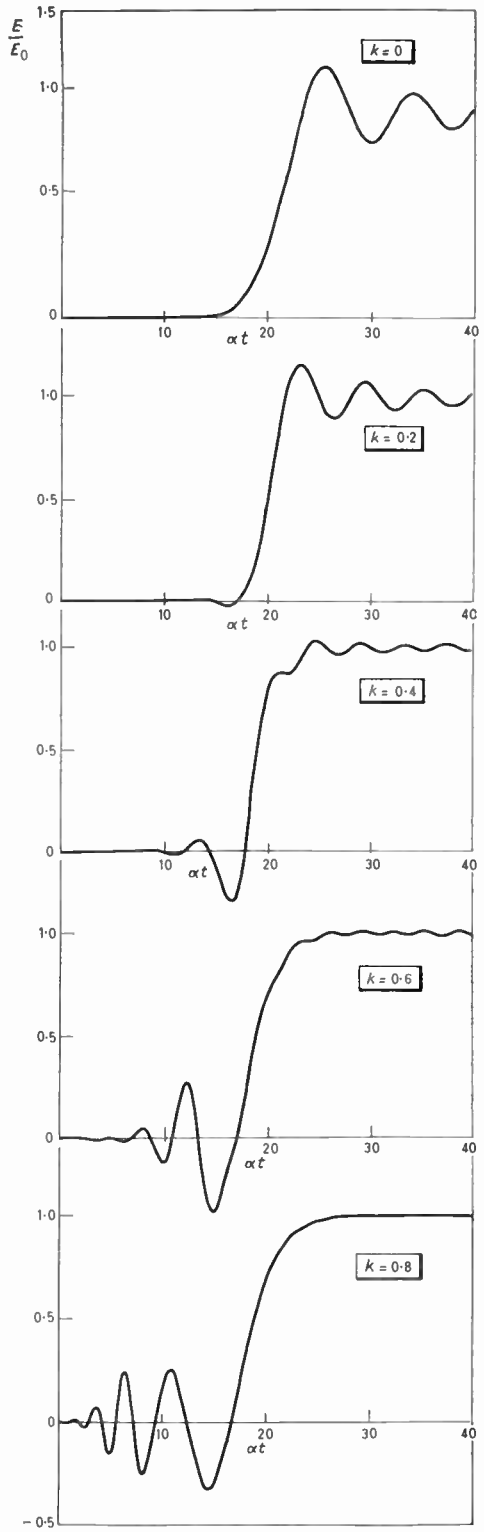

Fig. 7. Response of a ten-section *m*-derived delay line to unit step excitation with various degrees of coupling. Step generator of internal resistance equal to the characteristic impedance of the line (i.e. line terminated at both ends).

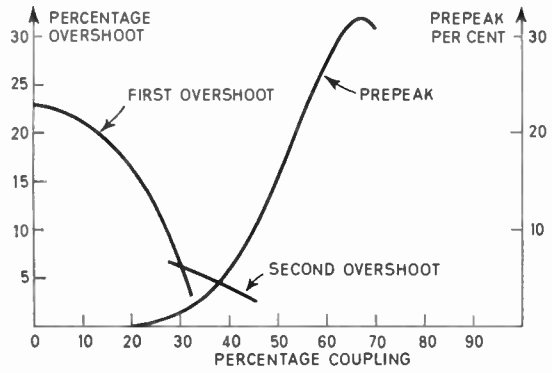

Fig. 8. Overshoot and prepeak of a ten-section *m*-derived delay line as a function of the degree of coupling between the sections. Line terminated at both ends (as in Fig. 7).

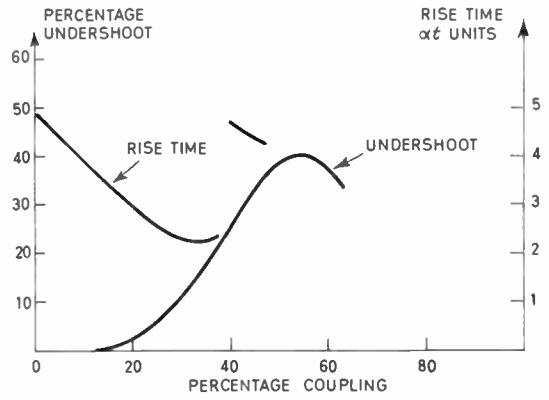

Fig. 9. Rise-time and undershoot of a ten-section *m*-derived delay line as a function of the degree of coupling. Line terminated at both ends (as in Fig. 7).

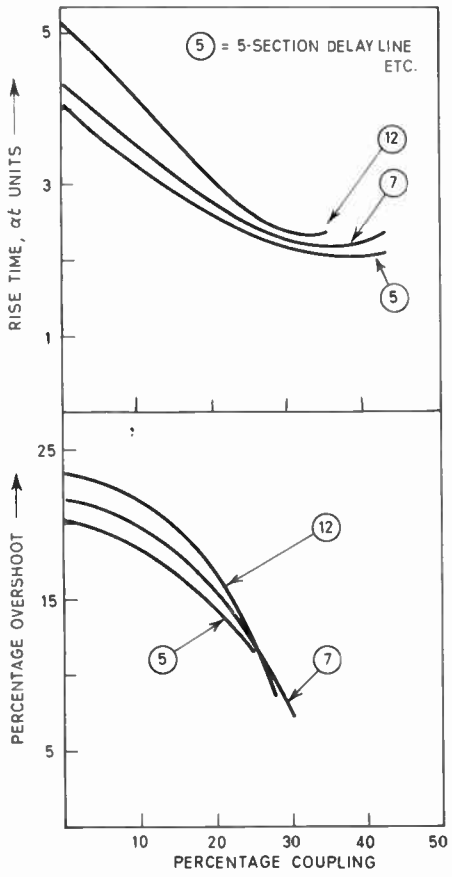

Fig. 10. Overshoot and rise-time of *m*-derived delay lines as a function of the degree of coupling. Various numbers of sections terminated at both ends.

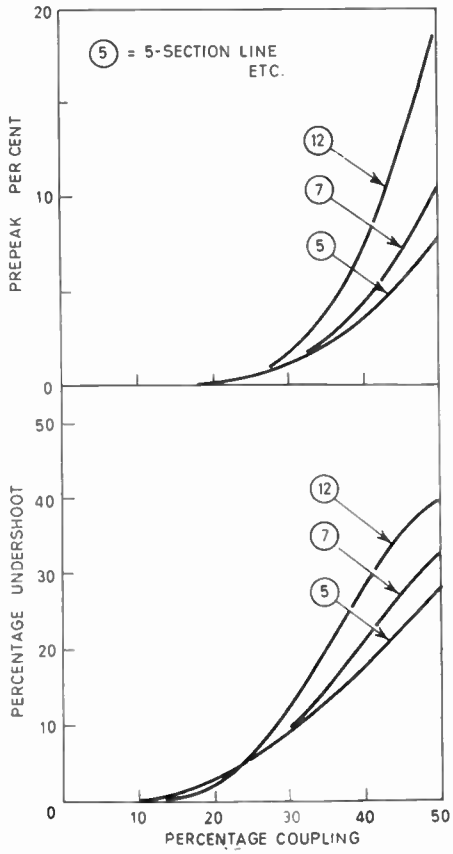

Fig. 11. Prepeak and undershoot of *m*-derived delay lines as a function of the degree of coupling. Various numbers of sections terminated at both ends.

The output of a ten-section delay line is shown graphically in Figs. 6 and 7. The results show that the coupling between the coils has two important beneficial effects. The overshoot and rise-time are both significantly reduced. However, new effects also appear which limit the maximum amount of coupling which may be used. Oscillations appear before the desired output, and their effect is first felt as a negative undershoot before the rise, and then also as a leading prepeak. These effects are graphically summarized

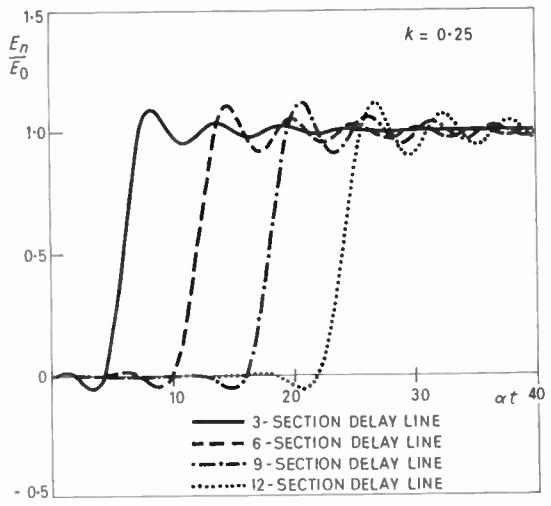

Fig. 12. Pulse response of *m*-derived delay lines for various numbers of sections terminated at both ends.

in Figs. 8 and 9. Near $k = 0{\cdot}4$ there is a sudden jump in the rise-time because of the wiggles which develop on the rising portion of the response. Indeed, the usual notion of rise-time is no longer useful in this region. The effect of variable coupling on delay lines with various number of sections is shown in Figs. 10 and 11.

From these figures it may be inferred that for delay lines with a large number of sections the useful coupling cannot be more than about $k = 0{\cdot}3$. Indeed, the optimum coupling used in practice is about 20 to 25%.

At $k = 0{\cdot}25$ the overshoot and rise-time are improved by about a factor of two. The leading pre-peak is then unimportant, being of the order of 1%. However, the undershoot is quite distinct since its amplitude is about −5% of the final output. This small undershoot is usually unimportant in practical network. Its presence reveals that positive coupling exists between the sections of the line.

The response of several delay lines with $k = 0{\cdot}25$ is shown in Fig. 12. These graphs are representative of the optimum results which may be achieved with this type of passive networks.

4. References

1. M. Avinor and Y. Partom, 'Pulse response of delay lines: Part 1—Constant-*k* delay lines', *The Radio and Electronic Engineer*, **29**, No. 5, pp. 271–6, May 1965.
2. A. Papoulis, 'A new method of inversion of Laplace transforms', *Quart. J. Appl. Maths*, **14**, p. 405, 1956–1957.

Manuscript first received by the Institution on 8th October 1965 and in revised form on 13th March 1966. (Paper No. 1071.)

Letter to the Editor

SIR,

The Use of S.I. Units

The recent article[1] on the use of these units contained no reference to the unit siemens (symbol S) as the replacement for the mho (reciprocal ohm), no doubt due to its omission from B.S.I. publication PD.5686. The latter omission was somewhat surprising as B.S. 1991: Part 6: 1963 had made reference, on page 28, to the adoption of the unit siemens and the symbol, S, by the I.E.C. and I.S.O.

It is most desirable that some lead be given by the Institution concerning the use of this unit, as it has already been widely adopted in North American publications. The American Institute of Electrical and Electronics Engineers have emphasized the importance of differentiating between the abbreviations ms and mS for millisecond and millisiemens respectively, by listing both of these units and their symbols in their Standard symbols for units.[2]

Yours faithfully,

G. MAY,
Associate

10 Bishops Road,
Farnham, Surrey.
16th August 1966.

[The reason for not including the siemens in the list of derived units was indeed its omission from the B.S.I. Publication PD.5686. The adoption of this unit by the I.E.C. and I.S.O. is definite and may well be accepted in subsequent B.S.I. Standards. There would not seem to be any good reason for continuing to use 'mho' and the typographically difficult ℧: certainly the necessary distinction between s for second and S for siemens is no more taxing to the reader, printer or typist than, say, m milli and M mega!

It is however interesting to note that the Institute of Electrical and Electronics Engineers has reconsidered this point, since in March of this year the use of mho rather than siemens was specifically stated to be preferred on the grounds that the former has been more widely used.[3]—F.W.S.]

References

1. *The Radio and Electronic Engineer*, **32**, No. 1, pp. 61–64, July 1966.
2. 'I.E.E.E. Standard Symbols for Units', *I.E.E.E. Transactions on Engineering Writing and Speech*, EWS-8, pp. 23–26, June 1965.
3. 'I.E.E.E. Recommended Practice for Units in Published Scientific and Technical Work', *I.E.E.E. Spectrum*, **3**, No. 3, pp. 169–73, March 1966.

U.D.C. 621.375.426

Gain and Stability of Tuned Transistor R.F. and I.F. Amplifiers

By

M. V. CALLENDAR, M.A., C.Eng.†

Summary: The high level of feedback in transistor r.f. amplifiers is well known, and neutralizing must normally be regarded as merely reducing (not eliminating) this feedback. Criteria are established for gain and stability, and also for distortion of the frequency-response curve. From the latter criterion, very simple practical design rules are arrived at: these are applicable to both single and multistage amplifiers, at any frequency. The results are summarized in a graph, which gives 'maximum useful gain' and design data on the preferred (fixed neutralized) design system: corresponding data for a non-neutralized system are also shown.

List of Principal Symbols

A_{pu} maximum unilateralized stage power gain ($= |y_f|^2/4g_i g_o$).

A_p insertion power (transducer) gain for a single stage amplifier.

A'_p insertion power gain of a (iterative) stage in a multistage amplifier chain.

S Stern[1] stability factor.

S_i Rollett[3] overall stability factor.

S_d response distortion criterion ($= S(1+\cos\phi_c)$).

y_f, y_r, y_i, y_o, g_i, g_o Y parameters of the transistor.

ϕ_f and ϕ_r are the phase angles of y_f and y_r.

K_r a feedback magnitude parameter, defined as $= \sqrt{\dfrac{|y_f y_r|}{g_i g_o}}$.

ϕ_c a feedback phase parameter, and $= \phi_f + \phi_r$.

Y_s and G_s refer to the input (current) generator.

Y_L and G_L refer to the output load.

$Y_1 = y_i + Y_s$ and $Y_2 = y_o + Y_L$.

$G_1 = g_i + G_s$ and $G_2 = g_o + G_L$.

(see Fig. 1.)

G_c the damping required to obtain stable gain (see Fig. 2).

G_{co} shunt conductance of a tuning coil (magnification Q_o).

σ interstage transformer ratio (usually $= \sqrt{g_i/g_o}$).

1. Introduction

The internal capacitive feedback in transistors used in tuned r.f. stages is usually more severe than with pentode valves. With present day transistors a study of the parameters shows that the level of feedback may sometimes approach a hundred times the minimum required to produce oscillation. A neutralizing circuit obviously cannot be expected to cope unaided with anything like this intensity of feedback; the ratio in which a typical neutralizing circuit (using a fixed capacitor) reduces the feedback is only 2 or 3 to 1 in the case of transistors which are on the limit for feedback capacitance.

It is usually necessary to use some other method (mismatching or damping), either to supplement the imperfect neutralizing or to replace it. This will call for knowledge of how much gain can be expected in practice from a given transistor, and what are the required values of source and load conductance, of damping conductance and of interstage transformer ratio.

Many writers have studied this problem. Of recent authors, W. Hetterscheid[5,6] gives a long and detailed analysis culminating in a number of computed response curves, but most other contributions[2,3,4] have been of mainly theoretical interest or have only dealt with particular aspects of the question.

One aspect barely touched on by most writers is the distortion of the response curve by feedback. In this paper, the curve-distortion is analysed, and a new design criterion is thus evolved. On the basis of this criterion, it has been found possible to arrive at *simple* design rules (for gain, damping, etc.) of *general* application at any frequency. By following these rules, stability, ease of alignment, and satisfactory response curve shape should be assured, though the stage gain will sometimes be a few decibels lower than that obtainable by other, more laborious, methods of design.

2. Power Gain

The gain and stability of two basic types of amplifier stage will now be considered.

† E. K. Cole Ltd., Southend-on-Sea, Essex.

2.1. *Single Stage Amplifier*

Figure 1 shows a single stage amplifier in Y parameter form. The problem is to determine how much the source and load conductances (G_s and G_L) have to be increased to maintain stability, and to calculate the power gain then obtained. The so-called 'transducer power gain' is here given by:

$$A_p = \frac{\text{power out to load}}{\text{available power from source}} = 4G_s G_L V_2^2/i_s^2 \quad \ldots\ldots(1)$$

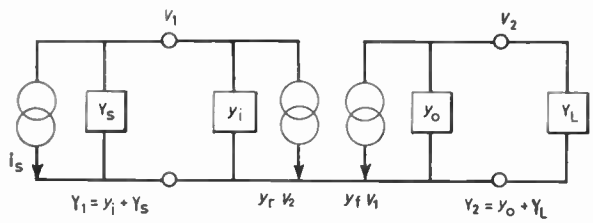

Fig. 1. Circuit for a single-stage amplifier (single-tuned), in Y parameter form. Y_s and Y_L each usually include a parallel tuned circuit.

This definition gives the gain obtained on inserting any amplifier between a source (Y_s) and load (Y_L), provided that we assume that the source is matched to the load by a transformer when the amplifier is not 'inserted'.

2.2. *Multistage Amplifier: Iterative Stage*

This may be formed from the single stage by inserting further ('iterative') stages between the source and the original stage. In a chain of such stages (assumed identical) as shown in Fig. 2, any one stage (here T_2 plus $T \times 2$ plus G_c) works out of a source (Y_s) equal to its own stage output admittance (Y_{out}) and works into a load (Y_L) equal to its own input admittance (Y_{in}). This condition is often

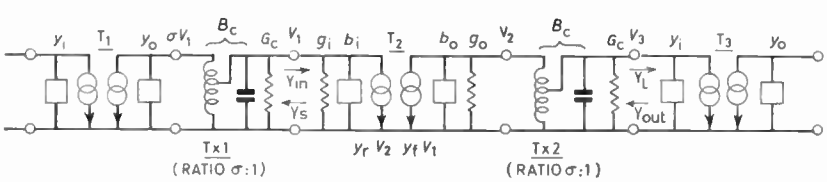

Fig. 2. Circuit showing iterative stages in a multistage amplifier.

referred to as 'iterative' matching, and the extra power gain obtained on inserting the stage in Fig. 2 is

$$A_p' = V_3^2/V_1^2 = |y_f|^2/\sigma^2(G_2)_0^2 \quad \ldots\ldots(2)$$

where $(G_2)_0$ is the total conductance measured at centre frequency f_0 between collector and earth, and σ is the transformer ratio.

(Note that this definition of insertion gain differs from the previous one in that the source is not in general matched to the load when the stage is not 'inserted'.)

An iterative chain may be stabilized either (i) by choosing the ratio (σ) of the interstage coupling to give a mismatch or (ii) by adding damping; this may take the form either of a shunt resistor, or of a deliberate use of the losses inherent in the coupling coils. It may be shown that both methods involve an equal loss of gain for a given improvement in stability. But the damping method will be assumed here, since it has the important advantage of minimizing the variation in gain and stability caused by differences in input (g_i) and output (g_o) conductance between different samples of a given transistor type. (The spread in these parameters is commonly from half to double the 'typical' value, and is often even more for g_o. Both g_i and g_o also vary with standing current, with a.g.c. bias, and with temperature.)

3. Criterion for Stability

The following well known formula applies to any device (e.g. a valve, or a transistor *at low frequencies*) having a resistive forward transadmittance (or 'slope') $y_f = g_f$ and a capacitive reverse admittance $y_r = -j\omega C_r$, using single tuned circuits (as Fig. 1)

$$S = 2G_1G_2/g_f\omega C_r \quad \ldots\ldots(3)$$

This S is called the 'stability factor' and is defined as the inverse of the maximum real loop gain. This means that if $S = 1$, the stage will *just* oscillate (at a particular frequency, f_s) but only when the tuning of both circuits (Y_1 and Y_2 in Fig. 1) is such as to give (at f_s) the zero loop-phase-shift required for oscillation. The criterion $S = 4$ has been conventionally accepted as offering a satisfactory margin.

With transistors in the common-emitter configuration the forward transadmittance shows a phase lag ϕ_{fe} which progressively increases with frequency, reaching around -90 deg at the highest useful frequencies. (The phase angle ϕ_{fb} in the common-base configuration is $= \phi_{fe} + 180°$.) This entirely alters the situation with regard to feedback: the following more general formula allowing for ϕ_f (and also for the fact that ϕ_r may not be exactly -90 deg) was originally devised by Stern[1]:

$$2G_1G_2 = |y_f y_r| . S(1 + \cos\phi_c) \quad \ldots\ldots(4)$$

Here $\phi_c = \phi_f + \phi_r$ and $\cos\phi_c$ is not far from zero at low frequencies, increasing with frequency to around -1 for common-emitter (or to around $+1$ in common base) at very high frequencies.

Recently another concept called 'invariant stability factor' has been introduced, and appears to have certain theoretical advantages. Of several possible defining formulae Rollet's 'overall stability factor' (S_i) seems to be the most relevant to practical design: this may be written as follows for comparison with the previous factor S

$$2G_1G_2 = |y_f y_r| . S_i \left(1 + \frac{1}{S_i} \cos \phi_c\right) \quad \text{.....(5)}$$

It is seen that the criterion for oscillation is unaltered (S and S_i here are both equal to 1), but that at higher levels of stability S_i tends towards $S(1+\cos \phi_c)$.

4. Response Curve Distortion

Stability, in the sense of freedom from oscillation, is not usually the only requirement for a satisfactory amplifier. A major effect of feedback, rather neglected by most writers, is to distort the shape of the frequency response curve, and at the same time to make alignment difficult.

Putting this into mathematical language, it can be shown that eqn. (4) above states that the maximum negative conductance $\frac{|y_f y_r|}{2G_2}(1+\cos \phi_c)$ reflected from the output circuit into the input circuit is equal to $1/S$ times the input circuit conductance G_1. But reflected reactance or positive conductance or both will be as important as negative conductance in causing curve distortion once we are well away from the condition ($S = 1$) for actual oscillation. So we should expect the curve distortion to be broadly proportional to the maximum reflected admittance, which is $\frac{|y_f|}{G_2} \times |y_r|$, giving us a new design equation

$$2G_1G_2 = |y_f y_r| . S_d \quad \text{.....(6)}$$

where $S_d = S(1+\cos \theta_c)$ may be called the 'curve distortion criterion'.

This rather general argument can be confirmed by a detailed analysis of the equations for the response curve, as given in Appendix 5. Curves calculated from these equations for a single stage with $Q_1 = Q_2$ are shown in Fig. 3: it is apparent that if we use $S = 4$ as our design criterion, we obtain very asymmetrical response curves at frequencies where $\cos \phi_c$ is around $-0{\cdot}5$ to $-0{\cdot}8$, and very flattened shapes when $\cos \phi_c$ approaches 1 (where the effective Q tends to zero). Setting $S_d = 4$ on the other hand keeps the asymmetry from rising excessively, and the damping is still reasonable (Q is decreased by 33%) even when $\cos \phi_c = -1$.

(In the case of common-base working, where $\cos \phi_c$ is positive, symmetry is never bad, but Q is doubled for $S_d = 4$ in the extreme case of $\cos \phi_c = +1$, so a higher value of S_d, say 8, may here be preferred.)

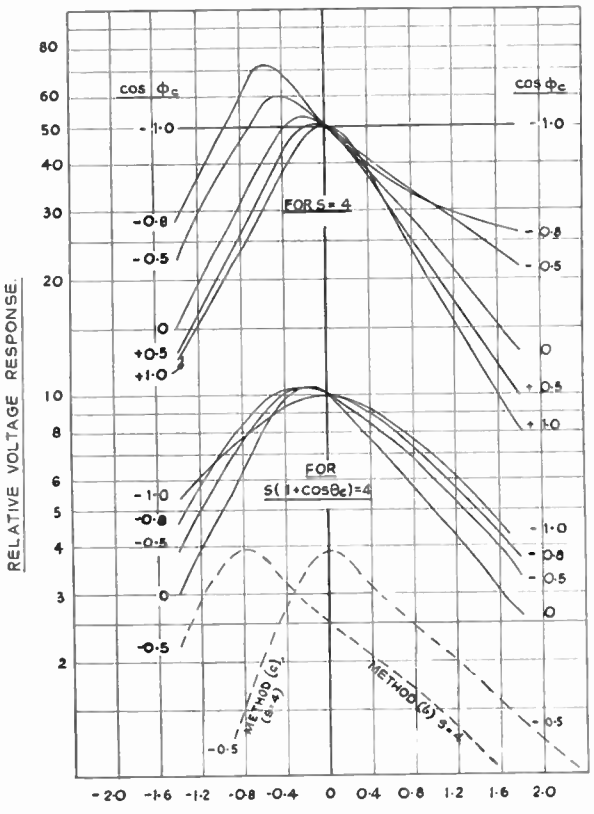

Fig. 3. Frequency response curves for a single stage without neutralizing with $\cos \phi_c$ as parameter. Method (*a*) tuning (see Appendix 3) is assumed except for the two curves marked. Shape of curves only (i.e. not gain) is shown.

The use of S_i instead of S_d as the criterion would give rather similar results for curve shape when S_i is greater than, say, 4 (compare eqns. (6) and (5)), but we shall base our designs upon S_d since this leads to much simpler formulae.

5. Gain Calculations: The General Case

We can, in principle, calculate the stability-limited gain or the curve-distortion-limited gain (for $S = 4$, or $S_i = 4$ or $S_d = 4$ for example), by putting eqns. (4), (5) or (6) into eqns. (1) and (2). But the matter is not, unfortunately, quite as simple as this.

One complication is that the correct method of tuning is by no means obvious and this greatly affects both gain and curve shape. Also, at very high frequencies the gain eventually ceases to be limited by stability, and must be calculated in a different way (see Appendix 1). Moreover, some writers have thought it necessary to give separate analyses for each case of one, two, three, etc., stages.

We thus end up with a dozen or more equations, most of which are very cumbersome. Moreover, each of these applies only under particular conditions which are not always definable in any easy way.

A few of the simpler formulae which can be derived are given in the Appendix, where tuning methods are also discussed.

6. A Practical Design Basis

From the practical standpoint it is essential to reduce this mass of information to something much simpler and more directly applicable. We start from the following basis:

(*a*) Neutralization, though far from a panacea, is still worthwhile, even when using a fixed capacitor.

(*b*) With $S_A = 4$ as our criterion, we have the simple design rule—make $G_1 G_2 = 2|y_f y_r|$. With fixed neutralizing, we can safely use only half this value for $G_1 G_2$ (see below).

(*c*) The most practical method of alignment is to apply a low impedance (say $< 10\,\Omega$) signal generator to the base of the last transistor, while tuning the output (collector) circuit, and to repeat this procedure at each stage in turn (see Appendix 3).

We shall now investigate the performance of amplifiers based on these premises. We start with common-emitter stages, using single tuned coupling circuits.

7. Stability and Curve-Distortion with Fixed Neutralization (by C_n) and $G_1 G_2 = |y_f y_r|$

With 'typical' transistors the stability is obviously good, and we need only examine the case of samples on the maximum and minimum limits for C_{bc}. If y'_r and ϕ'_c refer to the *nett* feedback admittance (with neutralizing circuit operative), we have

$$S'_d = \frac{2G_1 G_2}{|y_f y'_r|}$$

and

$$S'_i = \frac{2G_1 G_2}{|y_f y'_r|} - \cos\phi_c$$

Lowest and highest values for effective C_{bc} are best taken as 0·6 and 1·4 times the nominal value after including an allowance ($\pm 10\%$) for spreads in factors (e.g. C_n and coil ratios) outside the transistor. Then if $G_1 G_2 = |y_f y_r|$:

for high limit C_{bc}—

$\phi'_c = \phi_c$ approx. (since g_{re} is normally $\ll b_{re}$) and $|y'_r| = (1{\cdot}4 - 1{\cdot}0)|y_r|$. So $S'_d = 5$, and S'_i is always between 5 and 6;

for low limit C_{bc}—

$\phi'_c = (\pi + \phi_c)$ approx. and $|y'_r| = (1{\cdot}0 - 0{\cdot}6)|y_r|$.

So $S'_d = 5$ and S'_i is between 4 and 5.

However, if the ratio $G_1 G_2/|y_f y_r|$ were ever to be $< 0{\cdot}4$ of its nominal value (due to spreads in parameters not allowed for) instability could result at the higher frequencies, and here a slightly safer design is obtained by making the neutralizing $C_n = 90\%$ of its nominal value.

8. Gain with $G_1 G_2 = |y_f y_r|$, etc., as in Section 7

We may compare the stable gain to the well known 'maximum unilateralized gain' $A_{pu} = |y_f|^2/4g_i g_o$, and also to the ratio $|y_f|/|y_r|$.

The main complication in the general equations for gain arises from the terms representing input loading reflected from the output circuit (and vice versa), and the difficulties in tuning come from the same cause.

If we confine ourselves at first to 'typical' transistors, for which the neutralizing is not far from exact, $|y'_r| \simeq 0$ and the reflected loading is negligible, so that both the gain formulae and the alignment are much simplified. Putting K_r^2 for $|y_f y_r|/g_i g_o$, so that $G_1 G_2 = K_r^2 g_i g_o$ we have from eqns. (1) and (2):

Single Stage

$$A_p = \frac{4G_s G_L}{G_1^2 G_2^2} \cdot |y_f|^2 = \frac{4(K_r - 1)^2}{K_r^2} \cdot \left|\frac{y_f}{y_r}\right|$$

$$= \frac{16(K_r - 1)^2}{K_r^4} \cdot A_{pu} \qquad \text{.....(7)}$$

Required $G_s = (K_r - 1)g_i$ and $G_L = (K_r - 1)g_o$.

Iterative Stage

$$A'_p = \frac{|y_f|^2}{G_1 G_2} = \left|\frac{y_f}{y_r}\right| = \frac{4}{K_r^2} \cdot A_{pu} \qquad \text{.....(8)}$$

Required damping

$$G_c = (K_r - 2)g_i, \text{ with } \sigma^2 = g_i/g_o$$

These formulae are subject to the restriction that G_s/g_i and G_L/g_o should never be < 1 and that A_p can never exceed A_{pu}; these limitations are relevant only at the highest frequencies (see Sect. 9). Note that σ and the ratio of G_s to G_L are not critical provided $G_1 G_2$ is kept $= |y_f y_r|$. Where K_r is large, as at the lower frequencies, it is seen that

$$A_p = 4\left|\frac{y_f}{y_r}\right| = 4A'_p$$

The *average* gain taken over all samples will be as calculated above, but those with C_{bc} near the maximum limit will show a little less gain and those with C_{bc} near the minimum limit a little more gain. The difference between samples with limit and 'typical' values of C_{bc} will be about 3 dB (from eqns. (12) and (13) in the Appendix, since $G_1 G_2$ here $= 2{\cdot}5|y_f y'_r|$).

9. Gain and Stability at Very High Frequencies

Here ϕ_c in the common-emitter configuration is often near -180 deg (making $\cos\phi_c$ near -1) and $g_i g_o$ may be of the same order as $|y_f y_r|$, making K_r around 0·5 to 3 (see Fig. 5).

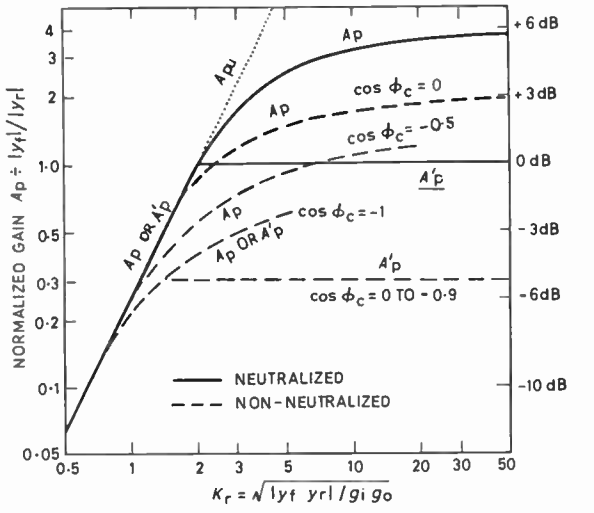

Fig. 5. Maximum useful gain (normalized) for a single stage (A_p) and for an iterative stage (A'_p). With corresponding design formulae. Solid lines are for the preferred (fixed) neutralized design as in Section 8, and dashed lines are for the non-neutralized design as in Section 10, (*a*). Curves apply for any value of $\cos\phi_c$ from 0 to -1 except where marked. Single-tuned coupling and 'normal' bandwidth are assumed.

Design formulae	Neut.	Non-neut.
Single stage: $G_s/g_i = G_L/g_o =$	$K_r - 1$	$1{\cdot}4K_r - 1$
Iterative: $\sigma^2 = g_i/g_o$ and $G_o/g_i =$	$K_r - 2$	$1{\cdot}4K_r - 2$
Subject to $G_s \ll g_i$, $G_L \ll g_o$, and $G_o \ll 0$		

At the highest frequencies, such that $4g_i g_o > |y_f y_r|$ (or $K_r < 2$), feedback is small and the simple conventional matching condition $G_s = g_i$ and $G_L = g_o$ (making $G_1 G_2 = 4g_i g_o$) is satisfactory. Gain will approximate to $A_{pu} = |y_f|^2/4g_i g_o$, and neutralizing may be omitted with little loss if $g_i g_o > |y_f y_r|$.

At rather lower frequencies where $4g_i g_o < |y_f y_r|$, our condition $G_1 G_2 = |y_f y_r|$ comes into operation and the gain is given by the formulae of the last section. We can show as follows that these formulae are still satisfactory even where $\cos\phi_c$ approaches -1. Under these conditions asymmetry is small, and the value of S_d becomes merely an indicator of the amount of increase in bandwidth due to feedback. For low limit C_{be} stability is passable ($S_i = 4$), with normal neutralizing. For high limit C_{bc}, the value of S_i is larger, and the loss of gain (due to the feedback being negative) is the only point requiring attention; from eqn. (10) in the Appendix, putting $G_1 G_2 = |y_f y_r| = 2{\cdot}5|y_f y'_r|$ and $\cos\phi_c = -1$, we see that the value of A_p is $1/(1+0{\cdot}4)^2$ times the value of A_p obtained when $|y_r| = 0$, so the gain for the limit sample is 3 dB less than for the typical sample, as in Section 8.

10. Alternative Bases for Design

Our preferred design basis, as described in the last four sections, appears to be satisfactory. But there are certain other possibilities.

(*a*) In some cases neutralization may not be practicable. With ultra-wide-band amplifiers, or those in which the frequency is variable, for example, the effective capacitance neutralized may vary too much over the band. A suitable design basis here is $G_1 G_2 = 2|y_f y_r|$ (see Appendix), giving $S_d = 4$ and $S_i = 4$ to 5 on a 'typical' sample transistor. As to gain (see Fig. 5) we usually lose 2 to 3 dB relative to the neutralized system owing to the lower $G_1 G_2$ product necessary for stability, plus a further 1 to 3 dB owing to damping imposed by the feedback. The stability margin and the response-distortion will be rather worse than with the neutralizing, except on samples near the minimum for C_{bc}.

(*b*) If we require to stretch the gain to the limit, we can obtain 2–3 dB more gain per stage by doubling our $G_1 G_2$ product; it can be shown that we should still just avoid oscillation. This applies both to our neutralized and non-neutralized designs, though in the former case it is here advisable to reduce the neutralizing to 80% of nominal when $\cos\phi_c > -0{\cdot}5$. But the variation in gain and the amount of curve-distortion with samples on the limit for C_{bc} will be roughly doubled, so that this scheme is usually inadvisable in amplifiers with a number of stages.

(*c*) We have hitherto assumed that no very large distortion of the response curve can be tolerated. But if this restriction is not necessary, we can obtain a further increase in gain *at the higher frequencies* where $\cos\phi_c$ is $-0{\cdot}5$ to $-0{\cdot}9$.

For example, if $\cos\phi_c = -0{\cdot}8$, we could use $G_1 G_2 = 0{\cdot}4|y_f y_r|$ without danger of oscillation ($S = 4$) even without neutralization. An increase in gain (as compared with $G_1 G_2 = 2|y_f y_r|$) of 7 dB is in principle realizable with exact neutralizing, except that of course the gain cannot exceed A_{pu}. In practice, the neutralizing, if fixed, must be restricted to 70% of the nominal value to avoid oscillation on low C_{bc} samples: the gain increase will be less than 7 dB and the gain spread between samples will usually be large.

Where feedback is strong, alignment is difficult, and the following method is suggested:

Tune each circuit in turn with an input of frequency f_0 with *both* adjacent circuits damped: the peak gain will then usually be found at a frequency (say $f_0+\Delta f$) slightly different from f_0, so the alignment is then repeated with input frequency $=f_0-\Delta f$ (see Appendix, Sect. 18.3): each circuit is finally readjusted slightly to maximize gain.

Except in the case of exact neutralization, gain calculations even for a single stage are very laborious under these conditions (see Appendix 6) and the results are sharply dependent upon the method of tuning assumed.

11. Common-Base Amplifiers

Common base stages give a lower gain than common-emitter stages except at the highest frequencies, but here, if neutralizing is not practicable, common-base stages have the advantage that their positive feedback gives them up to 6 dB higher gain than that obtained for common emitter (where the feedback is predominantly negative).

The Y parameter theory given above applies to common-base as well as to common-emitter. At frequencies high enough for the common-base circuit to be useful, $\cos\phi_c$ is order of $+1$, and $4g_{ib}g_{ob}$ is usually $> 2|y_f y_r|$; so simple matching ($G_s = g_{ib}$ and $G_L = g_{ob}$) is usable, and gain is one or two decibels in excess of A_{pu} (less coil losses). In general we should follow the rule of making $G_1 G_2 > 2|y_f y_r|$ (but see Sect. 4). Note that neutralizing here reduces gain slightly, especially in cases where $G_1 G_2$ is not much greater than $2|y_f y_r|$.

12. Double-Tuned or Band-Pass Coupling

With double-tuned coupling (see Fig. 4) the 'safe' design (giving S_i and $S_d \geqq 4$ for *any* tuning of the circuits) is to make G_{c1} and G_{c2} equal to the G_s and G_L calculated for single-tuned coupling. In this case the gain when limited by stability is 4 dB lower than for single circuit coupling in an iterative stage, or about 8 dB lower in a single stage (see Appendix 4). The same arguments apply whether the stage is (fixed) neutralized or not.

We can obtain a gain only slightly lower than for single-stage coupling while maintaining freedom from oscillation (even when detuned) and from excessive curve-distortion when correctly tuned up, by reducing G_{c1} and G_{c2} (see Appendix 4). But alignment is then considerably more difficult, and for a given criticality of adjustment, the gain with double tuning is undoubtedly appreciably less than with single tuning.

13. Coil Losses

Except at the highest frequencies (i.e. where $4g_i g_o > |y_f y_r|$ in our preferred design), damping (G_c) is required on all circuits except the first and last in any amplifier. This damping is conveniently supplied by connecting the rest of the circuit on to a tap at $1/n$ of the coil turns, such that $n^2 = Q_o \omega L G_c$.

So the A'_p calculated for iterative stages need not be reduced to allow for coil losses unless $G_c < 1/Q_o\omega L$ i.e. unless *either* (*a*) the band width is narrow, making $\omega L = 1/\omega C$ rather small *or* (*b*) the stage requires little damping (G_c small) as at very high frequencies.

But the single stage A_p will always be somewhat reduced. The fraction of the power lost in the coil conductance G_{co} will be $\frac{G_{co}}{G_w} = \frac{Q_w}{Q_o}$ in each circuit, so that the true gain will be $A_p \times \left(1 - \frac{Q_w}{Q_o}\right)^2$ where Q_w and G_w refer to the complete loaded circuit, and Q_o and G_{co} to the coil by itself. If a single stage has double-tuned circuits, the loss will still be $\left(1 - \frac{Q_w}{Q_o}\right)^2$ approximately in cases where the coils can supply G_{c1} and G_{c2}.

14. Bandwidth

In valve amplifiers, the concept of a constant (voltage gain) $\times$ (bandwidth) product simplifies design greatly. But with transistors the problem is less straightforward owing to their comparatively low input and output resistance, and to the greater difficulty of ensuring stability. The design of amplifiers with specialized coupling circuits is outside the scope of this paper, but a few general remarks can be made.

The highest gain—as calculated in Section 8 and as given in Table 1—is obtained at what we may call the 'normal' bandwidth. In this case each circuit has a 3 dB bandwidth $G_1/\omega C_1$ or $G_2/\omega C_2$ where G_1 and G_2 are our basic values, and the total tuning capacitance C_1 or C_2 arises mainly from the transistors,

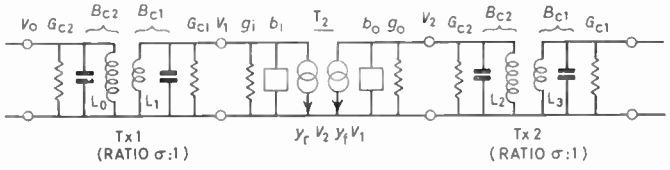

Fig. 4. Circuit showing double-tuned stages.

with no great additions. With transistors near the limit for C_{bc}, or with the non-neutralized design of Section 10, (*a*), some variation in bandwidth and deviation from the normal response shape must be expected, and with the unrestricted design of Section 10, (*c*) these variations may be very large.

For *wider bandwidth*, we merely increase the total conductances G_1, G_2 (i.e. we increase G_s, G_L and G_c) giving a proportional decrease in stage-gain (as with valve amplifiers) and also an increase in the value of S_i and S_d. Other techniques such as staggering or multiple circuits, can of course also be used.

For a *narrower bandwidth* we must add capacitances, keeping to our original values of G_1, G_2, etc. Here again we suffer a progressive, though at first slow, loss of gain, due to increase in power lost in the coils (Sect. 13).

It is interesting to note that the transistor giving most gain at 'normal' bandwidth is not necessarily best in other cases. As the bandwidth is increased, $|y_r|$ and g_i and g_o become rapidly less important, and eventually $|y_f|$, c_i and c_o are the only significant parameters. But if the bandwidth is made *narrower*, a low value of $g_i g_o$ becomes advantageous and A_{pu} eventually becomes a better 'figure of merit' than $|y_f|/|y_r|$. Indeed, $A_{pu}\left(1 - \frac{Q_w}{Q_o}\right)^2$ gives the 'narrow-band gain' under some conditions, but these conditions are unfortunately not very simply definable.

15. Practical Conclusions

It has been shown that a very simple design basis, viz. $G_1 G_2 = |y_f y_r|$ with fixed neutralizing or $G_1 G_2 = 2|y_f y_r|$ without, can be used to ensure good stability and response-curve shape in tuned transistor r.f. stages, at all frequencies. For the preferred (fixed neutralized) design, the formulae for gain and loading (Sect. 8) are also simple, and do not even require a knowledge of the phase angles ϕ_f and ϕ_r. It is only necessary to work out the value of $|y_f|/|y_r|$ and of $K_r = \sqrt{|y_f y_r|/g_i g_o}$.

The gain figures obtained on this design basis may be called the 'maximum useful power gain' (Ap or $A'p$) and Fig. 5 forms a convenient practical summary of the results, including design formulae. Figure 5 has been plotted from Sections 8 and 9, with Sections 18 and 19 for the non-neutralized circuit.

The value of G_s or G_L given in Fig. 5 may sometimes be unsuitable for special reasons, e.g. a lower value of G_s may be required to minimize noise factor on a first stage, and a low G_L is sometimes necessary to maximize voltage output from the last stage. In these cases the rule $G_1 G_2 = |y_f y_r|$ (or $= 2|y_f y_r|$ if not neutralized) should be applied directly, and gain will be little (less than 1 dB) below that given in Fig. 5, unless G_s or G_L differ greatly (over 3 times) from the values given there.

The ratio $|y_f|/|y_r|$ is a simple 'figure of merit' for comparing transistors for gain at 'normal' bandwidths and frequencies. But at very high frequencies the 'maximum unilateralized gain' A_{pu} must be used instead (see Fig. 5): at other frequencies a high value of A_{pu} is an indication of good gain at *narrow* bandwidth, and also permits some reduction in spread of gain between samples (since most of the damping can here be provided by the coils or by resistors).

Figures for 'maximum useful power gain', with fixed neutralizing in common-emitter, for a few well known types of transistors are given in Table 1. The columns headed A_p and A'_p were obtained from Fig. 5,

Table 1

Type	I_e mA	Frequency MHz	$\lvert y_f \rvert$ mmho	$\lvert y_r \rvert$ mmho	K_r	A_{pu} dB	A_p dB	A'_p dB
AF114	1	1·0	36	0·008	20	55	41·5	36
	1	10	32	0·08	9	39	31	26
	1	35	27	0·23	4·2	27	24·5	21
	1	100	18	0·5	1·35	14	14	14
AF114	4	35	60	0·23	2·2	25	24·5	24
AF181	3	35	85	0·08	3·3	35	33	30
AFZ12	1	35	35	0·13	24	46	30	24·5
	1	100	35	0·35	5·2	28	24	20
	1	200	35	0·65	2·7	20	19	17
SE3001	5	10	100	0·07	7·6	42	37	32
	5	35	70	0·20	5·4	34	30	26
	5	100	50	0·7	4·6	26	22	18·5
	5	200	43	1·4	3·8	20	18	15
	5	500	28	3·5	2·7	12	11	9

for a single stage and an iterative stage respectively. The gain of a complete N stage amplifier is given by $A_p + (N-1)A'_p$ (with A_p and A'_p in dB) to a good approximation.

In many cases stage gains will be rather lower than the 'maximum useful' figures given in Table 1. Thus:

(*a*) If neutralizing is not used, useful gain is up to 6 dB lower. (See Sect. 10 and Fig. 5.)

(*b*) Coil losses usually lower gain somewhat, especially at the higher frequencies, and this loss can be quite large at narrow bandwidths. (See Sects. 13 and 14.)

(*c*) With double-tuned couplings, gain is a little lower. (See Sect. 12.)

One great advantage of designing an amplifier for known performance, as above (instead of merely adding damping experimentally), is that it enables one to distinguish between feedback in the transistors themselves and that which arises from external causes—capacitive or magnetic coupling across one or several stages, common impedance in the emitter lead, insufficient d.c. coupling, etc. If, when using the calculated loading values (G_L, G_s, G_c), the amplifier lacks stability or does not give the calculated gain, the blame can be firmly placed on the external feedback paths, and measures taken accordingly.

16. Acknowledgment

The author desires to thank Messrs. E. K. Cole Limited for permission to publish this paper which is based on work done in their laboratories.

17. References

Note: These are selected from a much larger number available.

1. A. P. Stern, C. A. Aldridge and W. F. Chow, 'Internal feedback and neutralization of transistor amplifiers', *Proc. Inst. Radio Engrs*, **43**, pp. 838–47, July 1955.
2. W. W. Gärtner (correspondence), *Trans. I.R.E. on Circuit Theory*, **CT-5**, No. 4, December 1958.
3. J. M. Rollet, 'Stability and power-gain invariants of linear twoports', *Trans. I.R.E. on Circuit Theory*, **CT-9**, No. 1, pp. 29–32, March 1962.
4. S. Venkateswaran, 'An invariant stability factor and its significance', *Proc. Instn Elect. Engrs*, **109C**, pp. 98–102, 1962. (I.E.E. Monograph No. 468E, September 1961.)
5. W. Hetterscheid, 'Transistor i.f. amplifiers', *Electronic Applications*, **21**, No. 4, pp. 141–169, 1961.
6. W. Hetterscheid, 'Transistor Bandpass Amplifiers', (Cleaver Hume Press, London, 1964).

18. Appendix 1:
Analysis for a Stage with Feedback

Gain for a single stage with feedback, as Fig. 1. We have from eqn. (1)

$$A_p = \frac{4G_sG_L|y_f|^2}{|Y_2|^2\left|Y_1 - \frac{y_f y_r}{Y_2}\right|^2} \qquad \text{.....(9)}$$

When B_2 is made $= 0$ (i.e. Y_2 is tuned with Y_1 damped) and Y_1 is then tuned for maximum gain, we obtain at centre frequency f_0:

$$A_p = \frac{4G_sG_L|y_f|^2}{(G_1G_2 - |y_f y_r|\cos\phi_c)^2} \qquad \text{.....(10)}$$

(Note that other methods of tuning are possible: see Appendix 3 and Appendix 6.)

The maximum gain, without regard to stability, is obtainable by differentiating eqn. (10); this gives

$$\frac{G_s}{g_i} = \frac{G_L}{g_o} = \sqrt{1 - K_r^2\cos\phi_c}$$

and then

$$A_p = \frac{2A_{pu}}{1 - 0{\cdot}5K_r^2\cos\phi_c + \sqrt{1 - K_r^2\cos\phi_c}} \qquad \text{.....(11)}$$

Values calculated from eqn. (11) are shown in Fig. 5 as A_p for $\cos\phi_c = -1$.

But either stability or response-symmetry considerations or both together will (except where $\cos\phi_c$ is near -1) usually limit the gain to a value which is lower than the above A_p and can be evaluated by putting eqns. (4), (5) or (6) into eqn. (10). The simplest and most useful of these equations is the following, obtained by putting

$$S_d = S(1 + \cos\phi_c) = 4$$

into eqn. (10). Then

$$G_1G_2 = 2|y_f y_r|$$

giving

$$G_s/g_i = G_L/g_o = (1{\cdot}4K_r - 1)$$

and

$$A_p = X.|y_f|/|y_r| = A_{pu}.4X/K_r^2 \qquad \text{.....(12)}$$

where the factor

$$X = 2\left[\frac{1 - 0{\cdot}7/K_r}{1 - 0{\cdot}5\cos\phi_c}\right]^2$$

approximates to 2 at low frequencies (K_r large and $\cos\phi_c$ near 0) and falls at higher frequencies, as shown in Fig. 5 for $\cos\phi_c = 0$ and $\cos\phi_c = -0{\cdot}5$.

19. Appendix 2: Gain for an Iterative Stage

We have at f_0 from eqn. (2), putting $\sigma^2 = g_i/g_o$ to maintain symmetry,

$$A'_p = \frac{|y_f|^2}{(G_1G_2)_o} \qquad \text{.....(13)}$$

$(G_1)_o$ and $(G_2)_o$, the values of G_1 and G_2 at f_0, are in general more positive than G_1 and G_2 as given at f_s by eqns. (4) to (6), owing to the change with frequency of the conductances reflected into G_1 and G_2 from the adjacent stages.

Where we can neglect the reflected conductances (as when S_i or S_d is large) we have from eqns. (6) and (13)

$$A_p' = \frac{2}{S_d} \cdot \left|\frac{y_f}{y_r}\right|$$

and for $S_d = 4$ a full calculation (unfortunately too long to reproduce here) shows the loss due to reflected conductance to be 1·5 to 2·5 dB giving

$$A_p' \simeq 0{\cdot}3\left|\frac{y_f}{y_r}\right| \qquad \text{.....(14)}$$

Also

$$G_c = G_1 - 2g_i \simeq (1{\cdot}4K_r - 2)g_i$$

At a very high frequency where K_r has fallen to 1·4, it is seen that G_c becomes zero, and A_p' then becomes $= A_{pu}$ (see Fig. 5). The constants in these formulae vary less than ±10% over the range of cos.ϕ_c from 0 to −0·9, but when cos ϕ_c is near −1 a little more gain, approaching the line for A_p at cos $\phi_c = -1$ in Fig. 5, may be obtained with $G_c = 0$. (With actual transistors K_r rarely exceeds 3 when cos $\phi_c \simeq -1$.)

The gain is lower (except for cos $\phi_c = -1$) than for the single stage mainly owing to the loss in the damping resistor G_c.

20. Appendix 3: Tuning Procedure

The method of tuning assumed in Appendix 1 seems the most practical: in a multi-stage amplifier, each stage is merely tuned in turn (starting from the last) with a low impedance (say 10 Ω) generator connected to the base (or emitter). An alternative is to tune each circuit with *all* the others damped (e.g. making $B_1 = 0$ as well as $B_2 = 0$ in Fig. 1). This gives slightly different equations which show a little lower gain and usually more asymmetry, as typified by the curve marked Method (*b*) in Fig. 3. If we then retune all circuits to bring the peak to the centre (desired) frequency f_0, we obtain a little more gain (shown as Method (*c*) in Fig. 3): indeed, if $Q_1 = Q_2$ = etc., this seems to be the maximum possible gain, but the resultant curve is very lopsided except near f_0. The difference in gain between Methods (*b*) and (*c*) and Method (*a*) is normally < 1 dB under our condition $S_d = 4$ (or $S_i = 4$). But under conditions of lower stability as Section 10, (*a*), the differences in gain and curve shape become rapidly larger (see Appendix 6), and the 'best' method of tuning becomes more uncertain.

21. Appendix 4: Double-tuned or Bandpass Coupling (Fig. 4)

When a double-tuned coupling is used, the only really safe design basis (corresponding to that used with single circuits) is to postulate that the stage should exhibit the specified stability (e.g. $S_d = 4$) for any tuning of the four circuits L_0, L_1, L_2, L_3. This is equivalent to stating that the stage, consisting of $(L_1 + G_{c1} + T_2 + G_{c2} + L_2)$ should be fully stable on its own, so that the shunts G_{c1} and G_{c2} (which include the loss in the tuned circuits) must be equal to the correct source (G_s) and load (G_L) impedances evaluated in Appendix 1. The next stage ($L_3 + T_3$ + etc.) is then coupled in: so the total conductance $(G_2)_o$ at the collector of T_2 at f_0 is $2(G_{c2} + g_{22})$ or double that for single tuned (assuming critical coupling), and likewise for $(G_1)_o$. The fed-back conductance is also halved, however, and a full calculation for $S_d = 4$ shows that the gain for either a single or an iterative stage is 4 dB lower than the A_p' given in Section 2 for an iterative stage with single-tuned coupling, assuming cos ϕ_c between 0 and −0·9. This loss is due to the extra damping resistors (two extra on a single stage), and so disappears progressively at very high frequencies (for $K_r < 1{\cdot}4$).

The above loss relative to single tuned coupling can be roughly halved by making

$$\frac{G_{c1}}{g_i} = \frac{G_{c2}}{g_o} = (K_r - 1)$$

instead of $= (1{\cdot}4K_r - 1)$. We should still be just free from actual oscillation for any tuning of the circuits ($S > 2$) and curve-distortion should be fair (depending somewhat upon coupling factor), but alignment will be more critical than with the 'safe' design.

For a very detailed analysis of double-tuned coupling, together with calculated curves, reference should be made to Hetterscheid.[5,6]

Additional (undamped) circuits may be inserted in the couplings without, in principle, altering the gain, but in practice some allowance must be made for loss in the added circuits, as in normal filter practice.

22. Appendix 5: Response Curve Distortion

If in Fig. 1 the input circuit Y_1 is tuned to f_1 (i.e. $B_1 = 0$ at f_1) and Y_2 is tuned to f_2, we can write

$$F_1 = 2Q_1(f - f_1)/f_1 \quad \text{and} \quad F_2 = 2Q_2(f - f_2)/f_2$$

giving

$$B_1 = G_1 F_1 \quad \text{and} \quad B_2 = G_2 F_2$$

Then we can reformulate eqn. (9) in a form suitable for calculating the shape of the response curve:

$$A_p \propto |Y_1 Y_2 - y_f y_r|^{-2}$$

$$\propto \left|(1 + jF_1)(1 + jF_2) - \frac{y_f y_r}{G_1 G_2}\right|^{-2}$$

$$\propto [(1 + P - F_1 F_2)^2 + (F_1 + F_2 + T)^2]^{-1} \text{.....(15)}$$

When the gain is stability limited, we have

$$2G_1 G_2 = |y_f y_r| S(1 + \cos \phi_c)$$

so that

$$P = \frac{-R_c(y_f y_r)}{G_1 G_2} = \frac{-2\cos\phi_c}{S(1+\cos\phi_c)}$$

and

$$T = \frac{-2\sin\phi_c}{S(1+\cos\phi_c)}$$

From inspection it seems that the distortion of the curve shape by feedback (i.e. by the terms P and T) is not a function of S alone, but is mainly controlled by $S(1+\cos\phi_c)$. This can be confirmed as follows by expanding eqn. (15), putting $Q_1 = Q_2$ to keep the equations reasonably simple.

If we tune by making $B_2 = 0$ and then adjusting B_1 for maximum gain, as Method (*a*), we have $F_1 = F_2 - T$ and then:

$$A_p \propto [F_2^4 - 2TF_2^3 + 2F_2^2(1-P+0{\cdot}5T^2) + \\ + 2TF_2(1+P) + (1+P)^2]^{-1} \quad \ldots\ldots(16)$$

If we make S constant, the coefficient of F_2 (which mainly determines the tilt on the curve) increases about 10 times as $\cos\phi_c$ increases from 0 to $-0{\cdot}8$, whereas for $S(1+\cos\phi_c)$ constant this coefficient remains within $\pm 15\%$ over this range of $\cos\phi_c$, and is only 36% lower at $\cos\phi_c = -0{\cdot}9$. Likewise the frequency-invariant term $(1+P)^2$ (which represents the damping) rises 9 times as $\cos\phi_c$ increases from 0 to $-0{\cdot}8$ if S is held equal to 4, whereas for

$$S(1+\cos\phi_c) = 4$$

the rise is only 2 times. These points are borne out by the curves of Fig. 3.

If on the other hand we tune by Method (*b*), we have $f_1 = f_2$ and hence $F_1 = F_2$ giving:

$$A_p \propto [F_2^4 + 2F_2^2(1-P) + 4TF_2 + (1+P)^2 + T^2]^{-1} \quad \ldots\ldots(17)$$

The distortion of the response-curve for method (*b*) is somewhat similar to that discussed above. For Method (*c*) the shape is as for Method (*b*), but the centre frequency is shifted, giving increased dissymmetry except near f_0. If Q_1 differs from Q_2, the curve distortion seems to be less than for $Q_1 = Q_2$ provided that with Method (*a*) tuning the circuit having a higher Q is damped when the lower Q circuit is tuned: this means that our criterion $S_d = 4$ can ensure a good curve shape under all conditions.

When $\cos\phi_c = -1$, $T = 0$ and $P = |y_f y_r|/G_1 G_2$ giving a symmetrical but often broad curve: e.g. if we make $G_1 G_2 = 2|y_f y_r|$ the effective Q is two-thirds of that without feedback.

23. Appendix 6: Comparison with Other Theories

An attempt has been made to check the simple theory (assuming Method (*a*) tuning) presented in this paper from the viewpoint of Rollet,[3] Venkateswaran[4] and other recent writers who assume the amplifier to be tuned for maximum possible gain (instead of by Method (*a*)).

We start from an equation first given by Gärtner[5] for a single stage using an unconditionally stable device: this equation shows a transducer gain equal to that given by our eqn. (11) for $\cos\phi_c = -1$ but 3 dB greater for $\cos\phi_c = 0$. (Actual transistors are not unconditionally stable when $\cos\phi_c$ is near zero.) When, as commonly in practice, the device is not unconditionally stable, the equations become very unwieldy: no satisfactory explicit solution for A_p as a function of S_i seems to have been published, though Hetterscheid[6] gives a graphical method for A_p in terms of other parameters. The present writer has calculated some provisional figures for A_p with $S_i = 4$ and these come out about 1 dB greater than those for $S_d = 4$ given in Appendix 1.

Once the point has been passed at which there is any possibility of oscillation (allowing for parameter variation), the response curve distortion (as represented, unfortunately only roughly, by S_d) seems the only important criterion, and the figures representing the various possible 'stability factors' (e.g. S or S_i) are of mainly academic interest.

This investigation indicates that under conditions where response-curve distortion is important, the simple theory, assuming Method (*a*) tuning, is more useful. But for accurate calculations under the unrestricted conditions envisaged in Section 10, (*c*), the use of the more sophisticated theory would be unavoidable, in spite of the much more involved calculations and the greater practical difficulty in alignment. Again, detailed and laborious methods like those of Hetterscheid[6] must be used if an exact design, yielding the maximum possible gain, is required for double-tuned couplings.

In this connection, it may be noted that various *simple* formulae which have been published from time to time $\left(\text{e.g. } A_p' = \left|\frac{y_f}{y_r}\right| \cdot \frac{2}{S(1+\cos\phi_c)}\right)$ have usually been arrived at by neglecting the reflected admittances, and are therefore liable to large errors except when $\cos\phi_c$ is near 0 or when S is high.

Manuscript first received by the Institution on 30th July 1965 and in final form on 29th December 1965. (Paper No. 1072.)

Communications and Broadcasting Developments in the Commonwealth

NEW TYPE OF TELEPHONE LINK FOR EAST AFRICA

The East African Posts & Telecommunications Administration has placed an order with The Marconi Company for a new type of low capacity, radio communications system, which will provide six telephone channels between Mwanza and Bukoba, some 112 miles apart across Lake Victoria.

This new system, which is known as 'Thin Line Tropospheric Scatter', can provide a limited number of very reliable telephone or telegraph circuits over distances of up to 200 miles without the need for repeater stations. The system is free from the interference and distortion which are normally associated with high frequency radio circuits, while the cost of the equipment is considerably below that of the high capacity tropospheric scatter systems which have formerly been used for this type of work.

Conventional tropospheric scatter systems provide extremely high quality communication channels for medium range circuits, up to about 300 miles, linking points which are well over the horizon, without the need for repeater stations. This is achieved by transmitting a narrow beam of radio energy almost parallel to the ground, and in the required direction. A small part of this energy will be scattered by discontinuities in the troposphere, the lower part of the atmosphere, and a proportion of this scattered energy can then be picked up by sensitive receivers located well over the horizon from the transmitting station.

This type of scatter system is not affected by the daily and seasonal variations which cause interference and fading in high frequency radio systems, and the absence of repeater stations has obvious geographical, political and economic advantages. Tropospheric scatter systems normally use very high power transmitters to provide as many as sixty high-grade telephone channels with a reliability of 99·9% or better.

The cost of this type of tropospheric scatter system is however comparatively high, and in many parts of the world can only be economically justified for high circuit capacity. The new 'thin line' tropospheric scatter system is designed to provide a much lower channel capacity at a considerably lower cost, by using low-power transmitters with a reduced bandwidth. The system reliability is reduced to some extent, but is still considerably better than that of a normal h.f. radio circuit. It also has the major advantage of avoiding the congestion of the h.f. bands.

The new equipment has a narrow-band solid-state transmitter with an output of 7 to 10 watts in the frequency range 790 to 960 MHz. The frequency is crystal controlled and sensitive solid-state low-noise receivers are employed with 30-ft diameter dish aerials at both the transmitting and receiving stations. Frequency diversity is used to provide additional reliability in the system, without increasing the transmitting power.

This new telephone route across Lake Victoria will replace an indirect land-line route. It will increase the number of channels available between these two important centres, and will provide a considerable improvement in the speech quality. These new circuits will be integrated into the normal E.A.P. & T. telephone network.

The equipment is currently being developed by The Marconi Company at their Communications Laboratories at Writtle, near Chelmsford.

RADIO FACTORY IN ZAMBIA

Zambia's first radio manufacturing plant was recently opened by the President, Dr. Kenneth Kaunda. It is situated on the outskirts of Livingstone and it will produce transistor radios and radiogramophones for the Zambian market and for export. The plant has cost approximately £250,000 and is owned by Supersonic Radio (Zambia) Ltd., a subsidiary of Standard Telephones and Cables Ltd., of London.

NEW TRANSMITTING STATION AND STUDIO CENTRE FOR RADIO HONG KONG

Between now and January 1968, Radio Hong Kong's medium wave transmitting station at Hung Hom, in Kowloon, will be moved to Smuggler's Ridge in the New Territories. The Hung Hom site is required for the development of access roads to the projected cross-harbour tunnel linking Kowloon and Hong Kong Island.

Cable and Wireless Ltd., which has carried out all the engineering work of Radio Hong Kong, is to supervise the work of specifying and installing the new transmitters and the mast radiator. Site preparation and installation of the equipment will take several months, but it is hoped that testing of the new station will begin in October 1967. The Colony generally should get better reception of Radio Hong Kong's programmes as a result of the scheme.

Work is also in hand on the detailed design of the new £375,000 Broadcasting House which, it was announced last year, is to be built for Radio Hong Kong at Pokfulam, on Hong Kong Island. Four floors, with a total area of 53,000 square feet, will provide a compact studio centre, and will enable Radio Hong Kong to transfer all of its existing facilities from Mercury House, which it has occupied for the past 16 years.

The lower ground floor will house engineering maintenance facilities, air-conditioning plant and electric substation, record library archives, covered parking, a staff canteen and rehearsal rooms. The main studio floor, immediately above, will have four continuity and two talks suites, four production studios, three narrator's studios, and a music studio of 2400 square feet, as well as the main control room, record libraries and reception. The two upper floors will provide office accommodation and facilities for ancillary programme services.

The new Singapore Television Centre.

Much of the equipment that will go into Broadcasting House will be transferred from existing studios, but additional equipment is in store, or is being designed and built in Radio Hong Kong's own workshops.

Construction of the new building is expected to begin in May 1967, and to be completed 15 months later. A further four months will be required for the installation of equipment, and the building should be ready for occupation by the end of 1968.

Pokfulam is not within sight of either the Mount Gough v.h.f. f.m. transmitting station or the new Smuggler's Ridge medium wave station, and special arrangements will be made to ensure that reliable communication is maintained with the two transmitting stations. A special pressurized cable will be laid from Pokfulam to Mount Gough to carry essential programme lines and this will be designed for maximum security against damage by rain or typhoons. The Smuggler's Ridge Station will use the transmissions from Mount Gough to drive the medium wave transmitters.

SINGAPORE TELEVISION CENTRE

The new Television Centre of the Republic of Singapore is now in the final stages of construction, as the above photograph, taken earlier this year, shows. The Centre is designed to provide eventually four television services; initially two channels are to be sent out. There are two 80 ft × 60 ft studios, each with fully motorized lighting hoists operating from a lighting grid height of 42 ft. These grids also carry cranes permitting rapid striking of scenery backings during transmission. Besides the two major studios, two 40 ft × 30 ft studios cater for news and general magazine programmes.

The whole station is air conditioned and maintained at an ambient temperature of 73°F; in addition all equipment is independently cooled by forced air to maintain satisfactory temperature and humidity conditions.

The central technical area comprises a control apparatus room which controls communications and houses pulse generation equipment. Surrounding this central room are four continuity suites, one for each channel and each having a presentation studio. The telecine suite comprises four vidicon channels each with two projectors and slide facilities which can be multiplexed optically. The video-tape equipment comprises two Ampex VR 1000 C and two Ampex VR 2000, all operating in full intersync on the station timing network. The Marconi Mk. IV Image Orthicon Camera is standard to all studios.

The building was designed by the Singapore Government, and including structure, air conditioning and electrical equipment cost £400,000. Two years eight months have elapsed from first drawings to 'on-air'. The main electronics contractor was The Marconi Company.

U.D.C. 621.372.831

The Theory of Reflections in a Tapered Waveguide

By

R. A. WALDRON,
M.A.(Cantab), C.Eng.
(*Member*)†

Summary: The tapered waveguide is regarded as a two-mode system, the modes being the forward and reverse waves. A reflection parameter can be defined, relating the amplitude of the reverse wave generated at a point in the taper to the amplitude of the forward wave at that point. The reflection parameter can be regarded as a coupling parameter between the two modes, and coupled-mode theory can be used to obtain the overall reflection coefficient of the taper. It is shown how to estimate the error in this result. The error may become large if the frequency approaches the cut-off frequency of a guide whose cross-section is the same as that of the narrow end of the taper. The estimation of the error can be used to show how close to cut-off the theory is valid. A detailed analysis is given for the case of a homogeneously-filled tapered waveguide, with perfectly conducting walls, working in the dominant mode. The design of reflectionless tapers is discussed in the light of the results obtained.

List of Principal Symbols

The following list covers symbols used throughout the paper, and symbols occurring in formulae which express results. Symbols which occur only in one part of the paper are defined where they appear.

z distance along the waveguide, measured from $z = 0$ at the beginning of the taper.

z_0 value of z at the end of the taper.

$a(z)$ tapered dimension of the waveguide cross-section.

$Z_0(z)$ characteristic impedance of a uniform waveguide having the same value of a as the actual taper at z.

$Z(z)$ impedance seen on looking into the guide at z.

$\rho(z)$ reflection parameter, such that the reflection coefficient of a length δz, at z, is $\rho\delta z$.

$\beta(z)$ phase constant.

λ_0 wavelength in an infinite extent of the medium filling the taper. λ_0 will often be the free-space wavelength.

J see eqn. (6).

Y see eqn. (7).

R overall reflection coefficient of the taper.

χ constant depending on the shape of the waveguide cross-section. See Section 4, after eqn. (20).

A see eqns. (29)

p order of the first non-vanishing derivative of Z_0, β, or a at $z = 0$ or $z = z_0$.

n order of the second non-vanishing derivative of Z_0, β, or a at $z = 0$ or $z = z_0$.

Subscripts 0, 1, with ρ, β, Y, a, denote the values of these quantities at $z = 0$, $z = z_0$, respectively.

Dashes (′) denote differentiation with respect to z.

1. Introduction

The problem of matching a waveguide of one size to a similar waveguide of another size is often solved by tapering. A linear taper gives a first-order reflection which is often objectionable, and various smoothed tapers, which give only second-order reflections, have been proposed. The problem then arises of calculating the overall reflection coefficient of the smoothed taper, and this has not previously been satisfactorily solved, although a number of authors have claimed to have solved it. Only first-order treatments have been given,[1-3] and these are approximate; what is sometimes given as the overall reflection coefficient is in fact related to the error in the first-order estimate; the true first-order contribution to the reflection of a smooth taper is zero, and in previous treatments the second-order contribution has not been considered. Some authors, too, appear to confuse the reflection parameter, ρ, at $z = 0$, with the overall reflection coefficient, R.

Higher-order contributions are considered below, and it will be seen that as long as the reflection is small the major contribution to it is due to the values of $d^p a/dz^p$ at the ends of the taper, where a is the tapered dimension, z is distance along the waveguide axis from the start of the taper, and the pth is the first

† The Marconi Company Ltd., Research Division, Great Baddow, Essex.

non-vanishing derivative of a at the ends of the taper. The coefficients of the derivatives contain $(1-\chi^2/a^2)$, where χ is the cut-off value of a, raised to a negative power which increases with the order of the derivative. As cut-off is approached, therefore, the reflection becomes larger, and when this happens the contributions to the overall reflection due to the higher-order derivatives become important. When these higher-order contributions become comparable with that due to the pth derivative, the theory breaks down; this represents the nearest approach to cut-off for which a solution to the problem is obtainable. Tang[1] gives a solution in the form of a series, but the sum of this series is only the first-order contribution to the reflection. In principle, this can be calculated as close to cut-off as desired by taking enough terms of the series; however, this would not give a reliable result because higher-order contributions to the overall reflection are not considered.

In the present work we shall consider a reflection parameter, $\rho = \rho(z)$, as a coupling parameter between the forward and reverse waves. In Section 2 coupled-mode theory will be used to calculate the overall reflection coefficient, R, in terms of ρ. ρ itself will be considered in Section 3 and hence an expression for the overall reflection coefficient will be obtained in terms of the phase constant $\beta = \beta(z)$, and of $Z_0 = Z_0(z)$, which is the characteristic impedance of a uniform waveguide for which the cross-section is the same as that of the taper at the point z. β and Z_0 will in general vary with z, but in the special case of a transmission line with varying dimensions and uniform dielectric, β will be a constant. In Section 4, ρ, and hence R, will be obtained in terms of a for a waveguide with homogeneous dielectric filling, and the error in R will be discussed in the light of this result. In Section 5 the principles involved in designing a taper will be discussed in terms of the results obtained in the earlier sections, and an actual design will be treated in Section 6. The conditions under which the theory is valid will be considered in Section 7.

2. The Reflection of the Taper as a Coupled-Wave Problem

Consider the element of the taper between z and $z+\delta z$ (see Fig. 1). Assume that the input of the forward wave at z has unit amplitude, and the input of the reverse wave at $z+\delta z$ has zero amplitude. Due to the form of the taper, there will be a reflection proportional, in the limit as $\delta z \to 0$, to δz. Let the magnitude of this reflection be $\rho\delta z$. Then ρ is a function of z which we call the reflection parameter.

Now consider the forward wave, 1, to have amplitude $U_1(z)$ at z, and the reverse wave, 2, to have amplitude $U_2(z)$ at z. Then we can write the standard coupled-wave equations:

$$\left.\begin{aligned} \partial U_1/\partial z &= -\mathrm{j}\beta U_1 + \rho^* U_2 \\ \partial U_2/\partial z &= \mathrm{j}\beta U_2 + \rho U_1 \end{aligned}\right\} \quad \ldots\ldots(1)$$

where β is the phase constant of mode 1 and $-\beta$ that of mode 2. Multiplying by δz and adding U_1 to both sides of the first of these and U_2 to both sides of the second, and writing $U+(\partial U/\partial z)\delta z = U+\delta U$, we obtain

$$\left.\begin{aligned} U_1(z+\delta z) &= (1-\mathrm{j}\beta\delta z)U_1 + \rho^*\delta z U_2 \\ U_2(z+\delta z) &= (1+\mathrm{j}\beta\delta z)U_2 + \rho\delta z U_1 \end{aligned}\right\} \quad \ldots\ldots(2)$$

The problem is now to calculate from eqn. (2) the values of U_1 and U_2 at the ends $z = 0$, $z = z_0$, of the taper. A general solution to this kind of problem has been obtained[4] with β_1 and β_2 as the phase constants of waves 1 and 2, and with ρ^* replaced by $-\rho^*$ because the waves considered were travelling in the same direction. Replacing β_1, β_2, and $-\rho^*$ in the solutions of Reference 4 by β, $-\beta$, and ρ^*, we obtain as the solution of eqns. (2)

$$\begin{bmatrix} U_1(z_0) \\ U_2(z_0) \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} U_1(0) \\ U_2(0) \end{bmatrix} \quad \ldots\ldots(3)$$

where in this case

$$\left.\begin{aligned} T_{11} &= \frac{1}{\psi}\{Y_1 Y_0 \mathrm{e}^{\mathrm{j}J} - \rho_1^*\rho_0 \mathrm{e}^{-\mathrm{j}J}\} \\ T_{12} &= \frac{\mathrm{j}}{\psi}\{\rho_0^* Y_1 \mathrm{e}^{\mathrm{j}J} - \rho_1^* Y_0 \mathrm{e}^{-\mathrm{j}J}\} \\ T_{21} &= \frac{\mathrm{j}}{\psi}\{\rho_1 Y_0 \mathrm{e}^{\mathrm{j}J} - \rho_0 Y_1 \mathrm{e}^{-\mathrm{j}J}\} \\ T_{22} &= \frac{-1}{\psi}\{\rho_1\rho_0^* \mathrm{e}^{\mathrm{j}J} - Y_0 Y_1 \mathrm{e}^{-\mathrm{j}J}\} \end{aligned}\right\} \quad \ldots\ldots(4)$$

$$\psi = \sqrt{(Y_0^2 - \rho_0\rho_0^*)(Y_1^2 - \rho_1\rho_1^*)} \quad \ldots\ldots(5)$$

$$J = \int_0^{z_0} \sqrt{\beta^2 - \rho\rho^*}\,\mathrm{d}z \quad \ldots\ldots(6)$$

$$Y = Y(z) = -\beta + \sqrt{\beta^2 - \rho\rho^*} \quad \ldots\ldots(7)$$

The subscripts 0 and 1 with Y and ρ denote the values of these quantities at $z = 0$ and $z = z_0$ respectively.

Now put $U_2(z_0) = 0$; this states that there is no reverse-wave input to the far end of the taper. Then from eqn. (3) the overall reflection coefficient of the taper is

$$R = \frac{-T_{21}}{T_{22}} = \frac{\mathrm{j}\{\rho_1 Y_0 \mathrm{e}^{\mathrm{j}J} - \rho_0 Y_1 \mathrm{e}^{-\mathrm{j}J}\}}{\{\rho_1\rho_0^* \mathrm{e}^{\mathrm{j}J} - Y_0 Y_1 \mathrm{e}^{-\mathrm{j}J}\}} \quad \ldots\ldots(8)$$

Since ρ is small, we can expand Y_0 and Y_1 in terms of $\rho\rho^*/\beta^2$. Hence

$$R \simeq \frac{\mathrm{j}\rho_0}{2\beta_0} - \frac{\rho_0}{\rho_0^*} \cdot \frac{\rho_1^*}{2\beta_1} \mathrm{e}^{-2\mathrm{j}J} \quad \ldots\ldots(9)$$

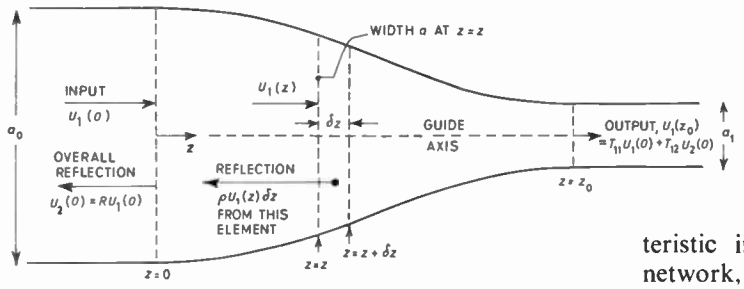

Fig. 1. Reflections in a tapered waveguide. The diagram shows, as an example, the broad dimension of a rectangular guide, the narrow dimension being constant and perpendicular to the plane of the diagram.

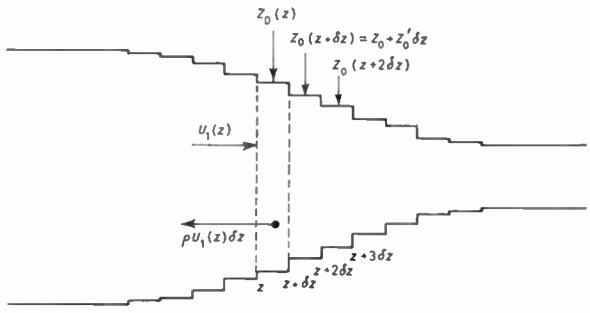

Fig. 2. Stepped approximation to the taper illustrated in Fig. 1.

To evaluate R it is necessary to know the first-order value of ρ and the value of β throughout the length of the taper; these are used in evaluating the integral J—although an approximate value that will usually be good enough can usually be obtained by neglecting $\rho\rho^*$ in comparison with β. It is also necessary, for the evaluation of R, to know the values of ρ and β at the ends of the taper.

It may appear at first sight that R can be made zero by making ρ_0 and ρ_1 zero. This is impossible, however, for ρ_0 and ρ_1 will be determined by the first non-vanishing derivative, at $z = 0$ and $z = z_0$, of the dimension a with respect to z, and if there is to be any taper at all there must be at least one derivative which is not zero at the ends—it may, of course, not be the same derivative at both ends. R can, however, be made zero as far as the pth-order reflection is concerned by making the two terms on the right-hand side of eqn. (9) equal and opposite, $d^p a/dz^p$ being the first non-vanishing derivative. We shall discuss this question further in Section 5.

3. Evaluation of the Reflection Parameter ρ

The tapered waveguide may be regarded as the limit of a stepped taper (Fig. 2). Consider the element δz from z to $z+\delta z$. The characteristic impedance of this is $Z_0(z)$ and at $z+\delta z$ it is joined to another element of characteristic impedance

$$Z_0(z+\delta z) = Z_0 + Z_0'\delta z$$

(see Fig. 3). Between the two is a four-terminal network due to the step. The element δz of characteristic impedance Z_0, together with the junction network, can be replaced by an equivalent line element of characteristic impedance $Z(z)$, different from Z_0 by an amount depending on the height of the step, i.e. on Z_0' (and higher derivatives) (Fig. 4). The reflection coefficient at the junction at $z+\delta z$ is then

$$\rho\delta z = \frac{(Z+\delta Z)-Z}{(Z+\delta Z)+Z} \qquad \text{.....(10)}$$

This ignores the effects of coupling into other modes, which can be expected to be negligible on the basis of experience of coupling problems in various kinds of deformed waveguide. We shall first calculate Z, which we shall call the apparent characteristic impedance of the tapered waveguide, and then use eqn. (10) to calculate the reflection parameter ρ.

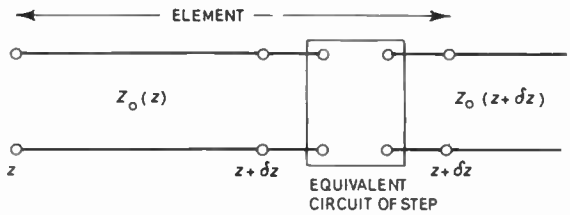

Fig. 3. Representation of the element from $z = z$ to $z = z + \delta z$ as an elementary length of uniform line loaded by a four-terminal network.

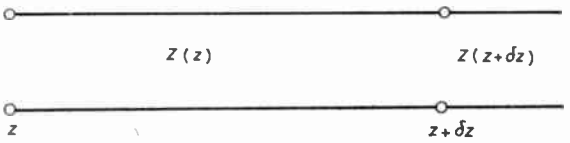

Fig. 4. Replacement of the element of Fig. 3 by an equivalent length of unloaded line. This is chosen to have an effective characteristic impedance $Z(z)$ such that the reflection coefficient, $\rho\delta z$, is unchanged.

3.1. *Apparent Characteristic Impedance of a Tapered Waveguide*

The relation of $Z(z)$ to $Z_0(z)$ may be calculated by regarding the element δz as having a characteristic impedance Z_0 and being connected to a line of characteristic impedance $Z(z+\delta z)$. This in its turn is connected to a line of characteristic impedance $Z(z+2\delta z)$, and so on, but the apparent characteristic impedance depends only on the nature of the element δz, not on what happens beyond it. The information on the

height of the step is contained in $Z(z+\delta z)$, but any further change in Z on going beyond $z+\delta z$ is irrelevant to the value of Z in the element under consideration. Looking at it another way, the impedance $Z(z+\delta z)$ will be seen as the input impedance at $z+\delta z$ only if the line beyond $z+\delta z$ is uniform with characteristic impedance $Z(z+\delta z)$. This model is used to calculate the apparent characteristic impedance $Z(z)$ at z.

In the limit as $\delta z \to 0$, the reflection from the element δz vanishes. We may therefore neglect it in the above model, and we can then say that the input impedance at z to the model is the apparent characteristic impedance at z, i.e.

$$Z(z) = Z_0(z)\left\{\frac{Z(z+\delta z)+jZ_0(z)\beta(z)\,\delta z}{Z_0(z)+jZ(z+\delta z)\beta(z)\,\delta z}\right\}$$

where $\tan \beta\delta z$ has been replaced by $\beta\delta z$. Write $Z(z+\delta z) = Z(z)+(\partial Z/\partial z)\delta z$ and expand to the first order in δz. Hence

$$Z(z) = Z(z)\left\{1+\left[\frac{\partial Z/\partial z}{Z}+\frac{jZ_0\beta}{Z}-\frac{jZ\beta}{Z_0}\right]\delta z\right\}$$

This equation balances only if the coefficient of δz vanishes, i.e.

$$\partial Z/\partial z - j\beta Z^2/Z_0 + j\beta Z_0 = 0 \qquad \text{.....(11)}$$

This equation, which is exact, can be solved by successive approximations. It is solved in Appendix 1 as far as the third approximation. Bolinder[2], Tang[1], and Walker and Wax[5] substituted for Z in terms of ρ and obtained solutions to the first order by neglecting terms in ρ^2 in the differential equation. In our case the solution is obtained directly in terms of Z. It is

$$Z = Z_0 - \frac{j}{2\beta}Z_0' - \frac{1}{4\beta^2}\left\{Z_0'' - \frac{Z_0'^2}{2Z_0} - \frac{Z_0'\beta'}{\beta}\right\} + \ldots \quad (12)$$

It will be noticed that the first approximation can be written

$$(-j/2\beta)^0 d^0Z_0/dz^0;$$

the second as

$$(-j/2\beta)dZ_0/dz;$$

the third has a first term

$$(-j/2\beta)^2\, d^2Z_0/dz^2.$$

It is interesting to speculate that the $(q+1)$th approximation may have a first term

$$(-j/2\beta)^q\, d^qZ_0/dz^q.$$

In fact, this turns out to be the case, as is shown in Appendix 2. The importance of this result will be seen in Section 3.2.

3.2. *The Reflection Parameter, ρ*

We can now use the expression (12) in eqn. (10) to evaluate ρ to the third order. We write first

$$Z = Z_0 + \Delta Z_0 \qquad \text{.....(13)}$$

where ΔZ_0 is the sum of the first-order and second-order terms (those in $1/\beta$ and $1/\beta^2$) on the right-hand side of eqn. (12). Substituting this into eqn. (10) we obtain

$$\rho\delta z = \frac{[(Z_0+\Delta Z_0)+(\partial Z_0/\partial z)\delta z + \\ +(\partial\Delta Z_0/\partial z)\delta z]-[Z_0+\Delta Z_0]}{[(Z_0+\Delta Z_0)+(\partial Z_0/\partial z)\delta z + \\ +(\partial\Delta Z_0/\partial z)\delta z]+[Z_0+\Delta Z_0]}$$

i.e.

$$\rho\delta z = \frac{[Z_0'+\partial\Delta Z_0/\partial z]\delta z}{2(Z_0+\Delta Z_0)+Z_0'\delta z+(\partial\Delta Z_0/\partial z)\delta z}$$

Cancelling the common factor δz, and then taking the limit as $\delta z \to 0$, this becomes

$$\rho = \frac{Z_0'+\partial\Delta Z_0/\partial z}{2(Z_0+\Delta Z_0)}$$

Expanding to the second order,

$$\rho = \frac{Z_0'}{2Z_0}\left\{1+\frac{1}{Z_0'}\frac{\partial\Delta Z_0}{\partial z}-\frac{\Delta Z_0}{Z_0}+\left(\frac{\Delta Z_0}{Z_0}\right)^2\right\}$$

We now substitute for ΔZ_0 the values of the appropriate terms of eqn. (12) and so obtain

$$\rho = \frac{Z_0'}{2Z_0}\left\{1+\left(\frac{-j}{2\beta}\right)\left[\frac{Z_0''}{Z_0'}-\frac{\beta'}{\beta}-\frac{Z_0'}{Z_0}\right]+ \right.$$
$$+\left(\frac{-j}{2\beta}\right)^2\left[\frac{Z_0'''}{Z_0'}-\frac{2Z_0''}{Z_0}+\frac{2Z_0'^2}{Z_0^2}-\frac{3Z_0''\beta'}{Z_0'\beta}- \right.$$
$$\left.\left.-\frac{\beta''}{\beta}+\frac{3\beta'^2}{\beta^2}+\frac{2Z_0'\beta'}{Z_0\beta}\right]\right\}\ldots\ldots \quad (14)$$

If Z_0' is non-zero, we may neglect all but the first-order term, and then

$$\rho \simeq Z_0'/2Z_0 \qquad \text{.....(15)}$$

This is the value of ρ to be used in evaluating the integral J, if the approximation is not made of neglecting $\rho\rho^*$ in comparison with β^2. In order to evaluate R (eqn. (9)) the values of ρ at the ends of the taper are required. If Z_0' and β' are zero at the ends, ρ is given by the second-order term in eqn. (14), i.e.

$$\rho \simeq \frac{-j}{2\beta}\cdot\frac{Z_0''}{2Z_0} \qquad \text{.....(16)}$$

If the second derivatives also vanish at the ends, we must go to the third derivative; then

$$\rho \simeq \left(\frac{-j}{2\beta}\right)^2\frac{Z_0'''}{2Z_0} \qquad \text{.....(17)}$$

Clearly, the value of ρ to be taken depends on the first non-vanishing derivative of Z_0 at the ends of the taper, and using the result obtained at the end of Section 3.1 for the first term of a higher-order contribution to Z, it is evident that if the first non-vanishing derivative of Z_0 is the pth, then

$$\rho \simeq \left(\frac{-j}{2\beta}\right)^{p-1}\frac{d^pZ_0/dz^p}{2Z_0} \qquad \text{.....(18)}$$

4. The Reflection of a Homogeneously-filled Metallic Tapered Waveguide

If Z_0 and β are known as functions of z, the theory developed so far can be applied to any waveguide or transmission line. The 'tapering' need not necessarily be a geometrical taper—it could be a progressive change of the properties of one of the media present in the guide. In order to be able to design a taper, it is necessary to be able to relate Z_0 and β to the geometry of the cross-section and to the properties of the media present. In general this can be a very difficult problem—if, indeed, the characteristic impedance can be said to be a meaningful concept in cases other than the simplest. However, an important class of waveguide tapers is that in which the guide is filled with a lossless homogeneous isotropic medium (probably air) and the walls, of metal, can be regarded as perfectly conducting. For this class, Z_0 and β, as functions of the geometry of the cross-section, are well known and we can treat such tapers in detail.

We have, for these waveguides,

$$Z_0 = \frac{N}{\sqrt{a^2-\chi^2}} \qquad \text{.....(19)}$$

$$\beta = \frac{2\pi}{\lambda_0}\sqrt{1-\chi^2/a^2} \qquad \text{.....(20)}$$

where λ_0 is the wavelength of waves in an infinite extent of the material filling the guide, $a = a(z)$ is a dimension of the cross-section, and N and χ are constants depending on the shape of the cross-section. For example, if a rectangular waveguide of height b and width a, filled with air, is tapered in the a dimension, we have $N = 240\pi b$, $\chi = \lambda_0/2$. For a guide of circular cross-section, a is the radius and $\chi = 1{\cdot}84118\lambda_0/2\pi$. In both these cases the dominant mode is assumed.

Substituting from eqns. (19) and (20) into eqn. (16) we obtain for the first-order reflection parameter

$$\rho = \frac{-\mathrm{d}a/\mathrm{d}z}{2a(1-\chi^2/a^2)} \qquad \text{.....(21)}$$

This is the value of ρ to be used in evaluating J. To evaluate ρ at the ends of the taper we substitute into eqn. (18) and obtain

$$\rho \simeq -\left(\frac{-\mathrm{j}\lambda_0}{4\pi\sqrt{1-\chi^2/a^2}}\right)^{p-1}\frac{\mathrm{d}^p a/\mathrm{d}z^p}{2a(1-\chi^2/a^2)} \qquad \text{.....(22)}$$

where $\mathrm{d}^p a/\mathrm{d}z^p$ is the first non-vanishing derivative of a at the ends of the taper.

These results are based on the assumption that the changes of Z_0 and β are the only factors contributing to the reflection, which will be the case if the taper is symmetrical. For an asymmetrical taper, the axis is curved and this curvature will also cause coupling. The value to be taken for ρ is then the sum of the value given in eqn. (21) or (22) and the value of the coupling coefficient due to curvature alone. In performing this addition, one must have regard to the relative phases of the contributions to ρ.

We can now insert the values of β_0 and β_1 from eqn. (20), and of ρ_0 and ρ_1 from eqn. (22), into eqn. (9). We notice that ρ is either real or imaginary, while β is real, since $a > \chi$, so that

$$R \simeq \frac{\mathrm{j}\rho_0}{2\beta_0} - \frac{\mathrm{j}\rho_1}{2\beta_1}\mathrm{e}^{-2\mathrm{j}J} \qquad \text{.....(23)}$$

Making the substitutions for ρ_0, ρ_1, β_0, and β_1, this becomes

$$R = \frac{-1}{2}\left(\frac{\mathrm{j}\lambda_0}{4\pi}\right)^p\left\{\frac{\mathrm{d}^p a_0/\mathrm{d}z^p}{a_0(1-\chi^2/a_0^2)^q} - \frac{\mathrm{e}^{-2\mathrm{j}J}\,\mathrm{d}^p a_1/\mathrm{d}z^p}{a_1(1-\chi^2/a_1^2)^q}\right\} \qquad \text{.....(24)}$$

where $q = 1+p/2$.

Equation (24) gives the overall reflection coefficient of the taper, to the lowest order that gives a non-zero result. The error in R can be estimated by considering the next-lowest order. In the nth-order contribution to ρ, there will be a term of the form of eqn. (22), with n instead of p. All the other terms, including those in $\mathrm{d}^p a/\mathrm{d}z^p$, will contain the 1st, 2nd, . . . $(p-1)$th, $(p+1)$th, . . . $(n-1)$th derivatives of a and so will vanish, unless $n = qp$ (q an integer), in which case there will be a term involving $(\mathrm{d}^p a/\mathrm{d}z^p)^q$ (see Appendix 3). The next contributions to ρ_0 and ρ_1 will therefore be the same as in eqn. (22), with p replaced by n.

We must also consider the effect of the second-order contribution to ρ on the integral J. If $|\rho^2|$ is neglected in comparison with β^2 in evaluating J, the error in J will be of order ρ^2/β^2, i.e. of the second order, and so will not contribute to the overall error if $n = p+1$. If $|\rho^2|$ is taken into consideration to the first order, the error is obtained by writing the integrand of J as

$$\sqrt{\beta^2-(\rho'+\mathrm{j}\rho'')(\rho'-\mathrm{j}\rho'')}$$

where ρ' and ρ'' are the first-order and second-order contributions to ρ. The terms in $\rho'\rho''$ cancel, and the error is due to the term in ρ''^2, which is of the fourth order. This term needs to be considered only if $n \geqslant p+3$.

5. Design Considerations

The aim in designing a taper is to minimize the reflection. This can be done by so choosing the function $a(z)$ that the two terms on the right-hand side of eqn. (9) cancel out. For the tapers considered in Section 4, this condition becomes, from eqn. (24),

$$\frac{d^p a_1/dz^p}{d^p a_0/dz^p} = \frac{a_1(1-\chi^2/a_1^2)^q}{a_0(1-\chi^2/a_0^2)^q} e^{2jJ} \qquad \text{.....(25)}$$

where again $q = 1+p/2$.

Equation (25) can evidently not be satisfied unless e^{2jJ} is real, i.e. unless $J = m\pi/2$, m being an integer. Further, the sign of the left-hand side will be that of $(-1)^{p+1}$ if the taper is monotonic, so that we require m odd if p is even, m even if p is odd. This indicates that if a design of order p is found to give an unsatisfactory reflection when the condition (25) is satisfied, the next consideration should be a design of order $p+2$ rather than $p+1$; then only the ends of the taper need be reshaped. If a change of one order is made, the value of J would be wrong and the whole taper would need redesigning.

Equation (25) expresses one condition for a minimum reflection. Others are that the values of a_0 and a_1 be what they are required to be, that J be $m\pi/2$, and that the derivatives of order lower than p should vanish at $z = 0$ and $z = z_0$. There are thus $4+2(p-1) = 2(p+1)$ conditions to be satisfied, so that the function $a(z)$ should contain $2(p+1)$ parameters which can be adjusted to meet the conditions. The design problem then reduces to the correct choice of these parameters. An example of this will be considered in Section 6.

J could be calculated from eqn. (6). There will, however, only be a second-order error, as pointed out above, if instead of eqn. (6) the formula

$$J = \int_0^{z_0} \beta \, dz \qquad \text{.....(26)}$$

is used. When eqn. (25) is satisfied, the overall reflection R will be given by an equation similar to eqn. (24) with p replaced by n, the order of the second non-vanishing derivative of a. If $n = p+1$, the second-order error in J will not matter. If $n = p+2$, it will matter, and this will complicate the calculation of the overall reflection.

The above conditions permit the minimization of R, by satisfying eqn. (25), at one frequency if the function $a(z)$ contains $2(p+1)$ parameters at the designer's disposal. Equation (25) can be satisfied at more than one frequency if the number of free parameters is increased, thus permitting broad-band design. The condition on J would also have to be satisfied at two frequencies, which means that for each extra design frequency two extra parameters are required.

The results of Gunston and Skwirzynski[6] indicate that, for a second-order reflection ($p = 2$), without satisfying eqn. (25), the reflection is already likely to be so small as to be difficult to measure. However, instead of reducing the overall reflection coefficient, it may be more desirable to design a shorter taper. That is, by relaxing somewhat the condition that Z_0, β, and a are slowly-varying quantities, the separate terms of eqn. (9) (or 23) might be made larger, and the overall reflection maintained fairly small by satisfying eqn. (25).

6. The Design of a Taper

We shall consider here the design of a taper, in which da/dz but not d^2a/dz^2 vanishes at $z = 0$ and $z = z_0$, with eqn. (25) satisfied at one value of frequency (i.e. one value of λ_0 and of χ, which depends on λ_0). The conditions to be observed are

(i) $a = a_0$ at $z = 0$.
(ii) $a = a_1$ at $z = z_0$.
(iii) $a' = 0$ at $z = 0$.
(iv) $a' = 0$ at $z = z_0$.
(v) Equation (25) is satisfied.
(vi) $J = m\pi/2$ (m odd).

Given a function $a(z)$ with enough parameters, the satisfaction of the first five conditions is a matter of elementary algebra and differential calculus, more or less complicated. The satisfaction of condition (vi) entails the evaluation of J, the integrand of which can be written as a function of a. It would be nice if a function $a(z)$ could be found such that the satisfaction of conditions (i) to (iv) led to simple expressions for the parameters of a, while at the same time J could be expressed in simple analytical form. Unfortunately, several guesses at $a(z)$ have failed to behave in this obliging fashion, so we shall use instead a function which makes conditions (i) to (v) easy to deal with, and let J take care of itself. It is

$$\frac{\chi^2}{a^2} = \frac{t(z^2-pz+q)}{z^2-rz+s} \qquad \text{.....(27)}$$

It is also convenient, for practical reasons, if the taper is monotonic. This is, in fact, found to be the case. It may be remarked that there are only five explicit parameters in this expression, instead of the six given by $2(p+1)$ with $p = 2$. The sixth parameter is z_0.

Applying conditions (i) to (v) to the above function, we obtain values for p, q, r, s, t, in terms of a_0 and a_1, which are given, and z_0. Hence

$$\frac{\chi^2}{a^2} = \frac{\chi^2}{a_0^2}\left\{\frac{z^2-2zz_0A+z_0^2A}{\alpha z^2-2zz_0A+z_0^2A}\right\} \qquad \text{.....(28)}$$

where

$$\left.\begin{aligned} A &= X/(1+X) \\ \alpha &= (X+a_1^2/a_0^2)/(1+X) \\ X &= \frac{a_1}{a_0}\left(\frac{1-\chi^2/a_1^2}{1-\chi^2/a_0^2}\right) \end{aligned}\right\} \qquad \text{.....(29)}$$

We now have to satisfy condition (vi). Substituting for χ^2/a^2, we obtain

$$J = \frac{2\pi z_0}{\lambda_0} \int_{u=0}^{1} \sqrt{\frac{Pu^2 - 2A(a_0^2-\chi^2)u + A(a_0^2-\chi^2)}{Qu^2 - 2Aa_0^2u + Aa_0^2}}\,\mathrm{d}u \qquad \ldots\ldots(30)$$

where

$$u = z/z_0$$
$$P = Q - \chi^2$$
$$Q = [a_1^2(1-A) + a_0^2 A]$$

The integrand is an elliptic integral; whether it is more convenient to evaluate it analytically and evaluate the result from tables, or to evaluate it by numerical integration, is left to the discretion of the designer. Notice that the integrand contains only the given quantities a_0, a_1, χ, and the dimensionless variable u. Thus the result of the integration will be a pure number which cannot be varied by changing parameter values. J is proportional to z_0. Having evaluated the integral, it is a simple matter to choose z_0 to satisfy condition (vi).

It may appear at first sight that J can be put equal to $\pi/2$, i.e. $m = 1$. However, this may lead to the variation of Z_0 and β, with z, being too rapid for the theory to be valid. The conditions on the rate of variation of Z_0 and β are discussed in Section 7; m must be taken sufficiently large to permit the satisfaction of these.

7. Limitations of the Theory

Before leaving this subject, we must discuss the conditions under which the theory that has been developed is valid. Firstly, the coupled-mode theory of Section 2 depends on the assumption that β and ρ are sufficiently slowly-varying quantities. The meaning of 'slowly-varying' in this context is examined in Reference 4; using the notation of the present paper, we require

$$\left.\begin{aligned} \mathrm{d}\beta/\mathrm{d}z &\ll 2\beta^2 \\ \mathrm{d}\rho/\mathrm{d}z &\ll 2\rho\beta \end{aligned}\right\} \qquad \ldots\ldots(31)$$

These conditions must hold over the length of the taper, except at isolated points. Thus, at the ends of a second- or higher-order taper, where ρ vanishes, the second of these conditions is violated, but since this violation occurs only over a minute length the theory is not invalidated. Also, if a approaches χ closely at the narrow end of the taper, i.e. if cut-off is approached closely, the first of the above conditions will be violated if a' is not sufficiently small. But if a' is not sufficiently small, the guide will only be able to have a value of a near to the cut-off value over a short length, and the theory of Section 2 is not invalidated. We may conclude that in all situations of practical interest the coupled-wave theory will be valid, although it would be wise to check in any particular case, in case some special feature is present which has not been thought of here. One should not be too pessimistic if it turns out that conditions (31) are violated over substantial lengths of the taper, provided that they are not too badly violated, because in order to be able to express sufficient conditions in a neat form, in reference (4), the conditions were made more stringent than is strictly necessary.

For the case of a homogeneously-filled tapered waveguide, as discussed in Sections 4 and 5, the conditions (31) become

$$\frac{\lambda_0 a'\chi^2}{4\pi(a^2-\chi^2)^{\frac{1}{2}}} \ll 1 \qquad \ldots\ldots(32)$$

and

$$\frac{\lambda_0}{4\pi(a^2-\chi^2)^{\frac{1}{2}}}\left\{\frac{a''}{a'}a(a^2-\chi^2) - a'(a^2+\chi^2)\right\} \ll 1$$

If the first of these is satisfied, and bearing in mind that a is of the same order as χ, we can write, instead of the second,

$$\frac{\lambda_0 a''a/a'}{4\pi(a^2-\chi^2)^{\frac{1}{2}}} \ll 1 \qquad \ldots\ldots(33)$$

Let us now see whether the conditions (32) and (33) are satisfied by the function considered in Section 6, i.e. equation (28). Condition (32) becomes

$$\frac{\lambda_0\chi^2}{4\pi} \cdot \frac{a_0^2 z_0 z(z-z_0)}{a(z^2-2Azz_0-Az_0^2)^2} \cdot A(1-\alpha) \ll (a^2-\chi^2)^{\frac{1}{2}}$$

If a_1 is near cut-off, X is small and A is of order X, i.e. of order $(1-\chi^2/a_1^2)^2$ (see eqns. (29)). If a_0 is near cut-off, X is large, and A and α are of order 1. $(1-\alpha)$ is then of order $1/X$, i.e. of order $(1-\chi^2/a_0^2)$. Near the ends, the above condition is satisfied if a_1 or a_0, as the case may be, is not too near χ. At the ends, either z or $z-z_0$ is zero, and the condition certainly holds. The condition (33) becomes too complicated to be worth writing down here, but it is satisfied by the function of equation (28).

In developing the theory, we have considered only the lowest-order contribution to ρ. This is a valid procedure only if the next-lowest-order contribution is much smaller than the lowest-order contribution. The contribution of order p is given by eqn. (22), and that of order $(p+1)$ is given by writing $(p+1)$ instead of p in eqn. (22). The ratio of the $(p+1)$th to the pth contribution is to be small, i.e.

$$\frac{\lambda_0 \mathrm{d}^{p+1}a/\mathrm{d}z^{p+1}}{4\pi\sqrt{1-\chi^2/a^2}\,.\,\mathrm{d}^p a/\mathrm{d}z^p} \ll 1 \qquad \ldots\ldots(34)$$

at the end of the taper where a is close to χ. This requirement places a limit on how close to cut-off it is possible to use the theory of this paper. If $\mathrm{d}^{p+1}a/\mathrm{d}z^{p+1}$ happens to be zero, it will be necessary

to go to a higher derivative for the second-lowest-order contribution to ρ.

In principle, the theory can be extended if it is desired to work closer to cut-off than the condition (34) permits. It is then necessary to take into consideration higher-order contributions to ρ. These are not given by eqn. (22), which applies only if the pth is the first non-vanishing derivative of a. If it were necessary to go beyond the third-order contribution to ρ, eqn. (13) would have to be solved to a higher order than the third which we have taken in this paper. The procedure may be tedious, but can be done in principle.

An absolute limit to the applicability of the theory is set by the accuracy with which Z_0 and β can be expressed in terms of the waveguide parameters. Near cut-off, even with a mathematically perfect guide, with perfectly conducting walls, Z_0 and β depend critically on a. Slight errors in the cross-sectional dimensions, imperfections in the practical realization of the function $a(z)$, the finite conductivity of the walls, corrosion of the walls, then become important and cannot be allowed for in the calculations.

For the taper considered in Section 6, condition (34) works out to

$$\frac{3\lambda_0}{2\pi z_0\sqrt{1-\chi^2/a^2}} \ll 1$$

i.e.

$$(f-f_c)/f \gg 9\lambda_0^2/8\pi^2 z_0^2$$

where f is the working frequency and f_c the cut-off frequency. Thus to work at a frequency $x\%$ above cut-off, we require a taper of length

$$z_0 \gg 15\lambda_0/\pi\sqrt{2x}$$

For a wavelength of 3 cm, this comes to $z_0 \gg 10/\sqrt{x}$ cm, say $z_0 = 50/\sqrt{x}$, i.e. some 50 cm for a frequency 1% above cut-off. Under these conditions it will be legitimate to calculate the performance, using only the lowest-order approximation to ρ.

With the above dimensions, conditions (32) and (33) are satisfied.

8. Acknowledgement

I wish to thank Mr. M. A. R. Gunston for commenting on the manuscript. Permission to publish this paper has been given by the Director of Research of The Marconi Company.

9. References

1. C. C. H. Tang, 'Non-uniform waveguide high-pass filters with extremely steep cut-off', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Techniques*, **MTT-12**, pp. 300–9, May 1964.
2. E. F. Bolinder, 'Fourier transforms in the theory of inhomogeneous transmission lines', *Trans. Roy. Inst. Tech., Stockholm*, No. 48, 1951.
3. N. H. Frank, 'Reflections from Sections of Tapered Transmission Lines and Wave Guides', RL Report 43-17, 6th January 1943 (PB 8058).
4. R. A. Waldron, 'The theory of coupled modes', *Quart. J. Mech. and Appl. Math.*, **18**, pp. 385–404, 1965.
5. L. R. Walker and N. Wax, 'Non-uniform transmission lines and reflection coefficients'. *J. Appl. Phys.*, **17**, pp. 1043–45, 1946.
6. M. A. R. Gunston and J. K. Skwirzynski, 'A broad-band low vswr H-plane waveguide taper', *Microwave Journal*, **5**, pp. 69–71, December 1962.

10. Appendix 1

Solution of Equation (11)

Equation (11) is a Riccati equation and can be solved by making the usual substitution to remove the term involving the square of the dependent variable —in this case, the term in Z^2. We write

$$Z = jy'Z_0/\beta y \qquad \text{.....(35)}$$

where $y' = dy/dz$. Equation (11) becomes

$$y'' + \left[\frac{Z_0'}{Z_0} - \frac{\beta'}{\beta}\right] y' + \beta^2 y = 0 \qquad \text{.....(36)}$$

Since Z_0 and β are slowly-varying functions of z, we can solve this by successive approximations.

First Approximation. We neglect the term in y', which has a first-order coefficient, and consider as first approximation the solution of the equation

$$y_1'' + \beta^2 y_1 = 0 \qquad \text{.....(37)}$$

If β were constant, the solution to this would be $y_1 = \exp\{\pm j\beta z\}$. Substituting this into eqn. (35), the + sign gives $Z \simeq -Z_0$, the − sign $Z \simeq Z_0$. The latter is what we expect, so we reject the + sign. Even so, we have not yet found the solution we require to eqn. (37), because β is not constant. By analogy with the solution for constant β, let us guess the solution $y_1 = \exp\{-j\int \beta\, dz\}$. Differentiating, we obtain

$$y_1' = -j\beta y_1 \qquad \text{.....(38)}$$

$$y_1'' = (-j\beta' - \beta^2) y_1 \qquad \text{.....(39)}$$

Substituting from eqn. (39) into eqn. (37), the zero-order terms balance, although the first-order terms do not. Therefore the function

$$y_1 = \exp\left\{-j\int \beta\, dz\right\} \qquad \text{.....(40)}$$

is a solution of eqn. (37) in the first approximation. It is evidently also the first approximation to the solution of eqn. (36).

Second Approximation. Write

$$y_2 = \eta y_1 \qquad \text{.....(41)}$$

as the second approximation. η will be a zero-order quantity, whose rate of variation with z will be a

first-order quantity, since if the variation were a zero-order quantity it would mean that y_1 is not the first approximation to the solution of eqn. (36) and when the substitution was made the zero-order terms would not have cancelled.

Differentiate eqn. (41) and substitute into eqn. (36). Then substitute from eqns. (38) and (39) for y_1' and y_1''. Hence

$$y_1\left\{\eta''-2j\beta\eta'+\left(\frac{Z_0'}{Z_0}-\frac{\beta'}{\beta}\right)\eta'-j\beta\frac{Z_0'}{Z_0}\eta\right\}=0$$

η', Z_0'/Z_0, and β'/β are first-order quantities; η'' is a second-order quantity. Considering only first-order quantities, we have

$$2\eta'+\eta Z_0'/Z_0=0$$

Integrating,

$$\log_e\eta=\frac{-1}{2}\log_e Z_0$$

i.e.

$$\eta=1/\sqrt{Z_0} \qquad \text{.....(42)}$$

Therefore

$$y_2=\frac{1}{\sqrt{Z_0}}\exp\left\{-j\int\beta\,dz\right\} \qquad \text{.....(43)}$$

Third Approximation. Write

$$y_3=\varphi y_2=\varphi\eta y_1 \qquad \text{.....(44)}$$

φ will be a zero-order quantity, φ' second-order, and φ'' third- or higher-order. Substituting this into eqn. (36), and substituting the functions y_1 and η and their derivatives, we obtain

$$\frac{\varphi'}{\varphi}=\frac{j}{4}\left\{\frac{Z_0''}{\beta Z_0}-\frac{Z_0'^2}{2\beta Z_0^2}-\frac{Z_0'\beta'}{Z_0\beta^2}\right\}$$

where the zero- and first-order terms have cancelled out, and higher terms than those of the second-order have been neglected; these higher-order terms will be taken into account in the higher-order approximations. From the above, we obtain

$$\frac{\varphi'}{\varphi}=\frac{j}{4}\frac{d}{dz}\left(\frac{Z_0'}{\beta Z_0}\right)+\frac{j}{8}\frac{Z_0'^2}{\beta Z_0^2}$$

Integrating,

$$\log_e\varphi=\frac{j}{4}\frac{Z_0'}{\beta Z_0}+\frac{j}{8}\int\frac{Z_0'^2\,dz}{\beta Z_0^2}$$

i.e.

$$\varphi=\exp\left\{\frac{jZ_0'}{4\beta Z_0}+\frac{j}{8}\int\frac{Z_0'^2}{\beta Z_0^2}dz\right\} \qquad \text{.....(45)}$$

Therefore

$$y_3=\frac{1}{\sqrt{Z_0}}\exp\left\{\frac{jZ_0'}{4\beta Z_0}-j\int[\beta-Z_0'^2/8\beta Z_0^2]\,dz\right\} \qquad \text{.....(46)}$$

To evaluate Z we require y_3'/y_3. This is

$$\frac{y_3'}{y_3}=\frac{\varphi'}{\varphi}+\frac{\eta'}{\eta}+\frac{y_1'}{y_1} \qquad \text{.....(47)}$$

and using the functions y_1, η, and φ that we have found, this gives

$$\frac{y_3'}{y_3}=-j\beta-\frac{Z_0'}{2Z_0}+\frac{j}{8}\left\{\frac{2Z_0''}{\beta Z_0}-\frac{Z_0'^2}{\beta Z_0^2}-\frac{2Z_0'\beta'}{Z_0\beta^2}\right\} \qquad \text{.....(48)}$$

Substituting this expression for y'/y into eqn. (35), we obtain

$$Z=Z_0-\frac{j}{2\beta}Z_0'-\frac{1}{4\beta^2}\left\{Z_0''-\frac{Z_0'^2}{2Z_0}-\frac{Z_0'\beta'}{\beta}\right\} \qquad \text{.....(49)}$$

Further approximations to the solution of eqn. (11) could be obtained, if required, by multiplying y_3 by yet another function, whose derivative would be a third-order quantity. The next approximation would require yet another factor, whose derivative would be a fourth-order quantity. In principle, this procedure could be continued indefinitely. As discussed in Section 7, however, it is not likely that higher-order approximations would be required.

11. Appendix 2

The Form of Z when the First $(p-1)$ Derivatives of Z_0 Vanish

Equation (12) is a third approximation to the solution of eqn. (11), and is good enough for the purposes of this paper if Z_0' and Z_0'' are non-zero. If they are zero, however, there is no distinction between this solution and the first-order approximation $Z=Z_0$, and to this order of accuracy ρ is zero. To obtain a value for ρ it is necessary to obtain a solution for Z which is different from Z_0, which will mean taking the $(p+1)$th approximation if the pth is the first non-vanishing derivative of Z_0. The form of eqn. (12) suggests that under these conditions the $(p+1)$th approximation may be

$$Z=Z_0+\left(\frac{-j}{2\beta}\right)^p\frac{d^pZ_0}{dz^p} \qquad \text{.....(50)}$$

and we can test this by substitution into eqn. (11).

Now, the 1st, 2nd, . . . $(p-1)$th derivatives will only vanish at isolated points ($z=0$, $z=z_0$), so we may not put them equal to zero before making the substitutions. They must therefore be considered in our trial solution, but the precise way in which they occur in the trial solution need not be known. We take the trial solution in the form

$$Z=Z_0+\left(\frac{-j}{2\beta}\right)Z_0'+\left(\frac{-j}{2\beta}\right)^2\left\{Z_0''-\frac{Z_0'^2}{2Z_0}-\frac{Z_0'\beta'}{\beta}\right\}+$$
$$+\ldots+\left(\frac{-j}{2\beta}\right)^{p-1}\left[\frac{d^{p-1}Z_0}{dz^{p-1}}+P_{p-1}\right]+$$
$$+\left(\frac{-j}{2\beta}\right)^p\left[\frac{d^pZ_0}{dz^p}+P_p\right]+\ldots\ldots \qquad (51)$$

where P_q consists of products of derivatives, all of order less than q, such that the sum of the orders in

each product term is q. For example, P_2 contains only first derivatives in products of two so that the orders add up to 2. Differentiating,

$$Z' = Z_0' + \left(\frac{-j}{2\beta}\right)[Z_0''+Q_2] + \left(\frac{-j}{2\beta}\right)^2[Z_0'''+Q_3] + \\ + \ldots + \left(\frac{-j}{2\beta}\right)^{p-1}\left[\frac{d^pZ_0}{dz^p} + Q_p\right] + \\ + \left(\frac{-j}{2\beta}\right)^p\left[\frac{d^{p+1}Z_0}{dz^{p+1}} + Q_{p+1}\right] + \ldots\ldots (52)$$

where Q_q, like P_q, is a term of order q made up of products of derivatives of orders less than q.

Now substitute for Z and Z' in eqn. (11).

$$\left\{Z_0' + \left(\frac{-j}{2\beta}\right)[Z_0''+Q_2] + \ldots + \right. \\ + \left(\frac{-j}{2\beta}\right)^{p-1}\left[\frac{d^pZ_0}{dz^p} + Q_p\right] + \\ \left. + \left(\frac{-j}{2\beta}\right)^p\left[\frac{d^{p+1}Z_0}{dz^{p+1}} + Q_{p+1}\right] + \ldots\right\} - \\ - \frac{j\beta}{Z_0}\left\{Z_0^2 + 2Z_0\left[\left(\frac{-j}{2\beta}\right)Z_0' + \ldots + \right.\right. \\ + \left(\frac{-j}{2\beta}\right)^{p-1}\left(\frac{d^{p-1}Z_0}{dz^{p-1}} + P_{p-1}\right) + \\ \left.\left. + \left(\frac{-j}{2\beta}\right)^p\left(\frac{d^pZ_0}{dz^p} + P_p\right) + \ldots\right] + \right. \\ \left. + \left[\left(\frac{-j}{2\beta}\right)Z_0' + \left(\frac{-j}{2\beta}\right)^2(Z_0'' + P_2) + \ldots\right]^2\right\} \\ + j\beta Z_0 = 0$$

Now put $Z_0' = Z_0'' = \ldots = d^{p-1}Z_0/dz^{p-1} = 0$. Then $P_1, P_2, \ldots P_p$ and $Q_1, Q_2, \ldots Q_p$ all vanish, and we are left with

$$\left\{\left(\frac{-j}{2\beta}\right)^{p-1}\left(\frac{d^pZ_0}{dz^p}\right) + \right. \\ \left. + \left(\frac{-j}{2\beta}\right)^p\left[\frac{d^{p+1}Z_0}{dz^{p+1}} + Q_{p+1}\right] + \ldots\right\} - \\ - \frac{j\beta}{Z_0}\left\{Z_0^2 + 2Z_0\left[\left(\frac{-j}{2\beta}\right)^p\left(\frac{d^pZ_0}{dz^p}\right) + \right.\right. \\ \left.\left. + \text{higher-order terms}\right] + \right. \\ \left. + [\text{terms of order } 2p \text{ and higher}]\right\} + j\beta Z_0 = 0$$

For a $(p+1)$th approximation to the solution only terms to order p need be considered. To this order the above equation is satisfied. Therefore at a point where the first $p-1$ derivatives of Z_0 vanish, Z takes the form of eqn. (50).

12. Appendix 3

The Next Non-Vanishing Derivative

If the next derivative of Z_0 after the pth which does not vanish at the ends is the nth, $P_{p+1}, P_{p+2}, \ldots P_n$ and $Q_{p+1}, Q_{p+2}, \ldots Q_n$ will vanish, if $n < 2p$. Then the $(n+1)$th-order solution for Z would be

$$Z = Z_0 + \left(\frac{-j}{2\beta}\right)^p\frac{d^pZ_0}{dz^p} + \left(\frac{-j}{2\beta}\right)^n\frac{d^nZ_0}{dz^n} \ldots\ldots (53)$$

This would also be the case if $2p < n < 3p$, $3p < n < 4p$, etc. But if $n = 2p$, $3p$, $4p$, or any other integer times p, P_n will contain a term involving the square of d^pZ_0/dz^p and the product of this with $d^p\beta/dz^p$. For example, if $p = 1$ and $n = 2$, these extra terms are given along with d^2Z_0/dz^2 in the bracket in eqn. (49). Thus eqn. (53) depends on n not being a multiple of p.

Manuscript first received by the Institution on 19th October 1965 and in final form on 24th February 1966. (Paper No. 1073.)

U.D.C.621.382.2/3:681.38

Graphical Derivation of Trajectories for Transistor–Tunnel Diode Logic Circuits

By

H. MADANI, B.Sc., Ph.D., C.Eng.†

Summary: A description is given of a graphical derivation of the switching trajectories of a transistor–tunnel diode circuit. With this method quicker solutions are achieved than when an analytical approach is used and the resultant accuracy compares favourably with that of solutions obtained using an analogue computer.

List of Symbols

$A_v = \omega_b/(p \times \omega_b)$; factor multiplying the input voltage, in the new common-collector equivalent circuit.
C_c transistor collector capacitance.
C_d tunnel diode capacitance.
C_e transistor emitter capacitance.
i_c current flowing through capacitance C_d.
i_d tunnel diode current.
i_e transistor emitter current.
I_0 constant current generated in Fig. 1.
$L_1 = \alpha_0 r_{bt}[1-(1-\alpha_0)\omega_n/\omega_b]/\omega_n$
L_d tunnel diode inductance.
L_0 externally added inductance in Fig. 1.
$L_t = L_d + L_0$
$r_1 = r_{bt}(1-\alpha_0)$.
$r_2 = \alpha_0 r_{bt}[1-(1-\alpha_0)\omega_n/\omega_b]/(1+\alpha_0\omega_n/\omega_b)$.
$r_{bb'}$ internal base resistance.
r_{bt} total resistance in the base circuit, including $r_{bb'}$ and any externally-added resistance, e.g. R_s.
r_d tunnel diode series resistance.
R_d voltage-dependent resistance of ideal tunnel diode.
r_e transistor emitter resistance.
$r_{ee} = r_e + r_{bt}(1-\alpha_0)$.
v_d voltage across tunnel diode, excluding the inductance L_d.
v_0 output voltage; for the tunnel diode, this refers to the total voltage across the tunnel diode, including the inductance L_d.
v_1 voltage across inductance L_t.
v_{in} input voltage.
v_4 voltage at emitter of TR1 in Fig. 1.
Z_t output impedance of the common-collector configuration.
α large-signal short-circuit current gain of the common base configuration.
α_0 value of α under d.c. conditions.
β large-signal short-circuit current gain of the common emitter configuration.
γ large-signal short-circuit current gain of the common-collector configuration.
ε energy gap.
$\omega_b = 1/r_{bt}C_c$.
$\omega_d = 1/R_dC_d$.
$\omega_n = 2\pi \times$ 3dB alpha cut-off frequency.

1. Introduction

The circuit under analysis is shown in Fig. 1, and was first described by Neff, Butler and Critchlow.[1] The purpose of analysing its switching performance was to provide information for a comparison between similar binary and ternary logic circuits,[2] and was part of a more general investigation into the possibility of using ternary logic in computers.

Different methods of analysing the switching performance of the circuits have been described.[3] The first method was to derive the fifth degree transfer function of the circuit and to solve it analytically. This method, though informative, tended to be laborious and of average accuracy, and was replaced by a digital computer solution, which led to more accurate results. A more versatile analogue computer solution, involving the use of the direct analogue

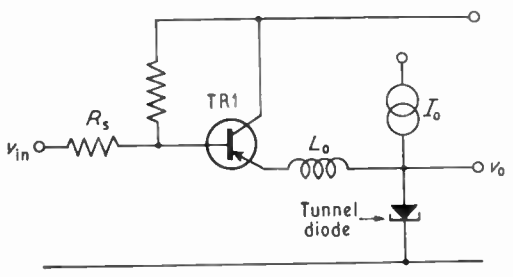

Fig. 1. Circuit for analysis.

† The American University of Beirut, Lebanon; the work described was carried out at the Queen's University of Belfast.

method, gave excellent results, and tended to yield the best combination of speed and accuracy. A twelve-segment function generator was used to generate the tunnel diode characteristic, while about twenty operational amplifiers were required for the remaining parts of the circuit. It was relatively easy to observe directly the effect of varying important circuit parameters on the performance of the circuit.

It is felt, however, that the circuit also lends itself to a graphical solution, which could be useful where there is no access to a digital or analogue computer, and where time and accuracy do not justify the lengthy hand-calculations of the analytical method mentioned above. One typical application of the graphical method is the need to augment certain results previously obtained with the help of a computer.

2. Previous Graphical Methods

Graphical methods have been described for the calculation of switching times of circuits,[4, 5] but of a simpler nature than the circuit under consideration here. Both references depend on the application of Preisman's method[6] to non-linear circuits, in which differentials are replaced by finite differences, thus allowing numerical integration. For example, an expression such as

$$v = \frac{1}{C}\int i_c . \mathrm{d}t \qquad \ldots\ldots(1)$$

may be approximated by

$$\Delta v = \frac{1}{C} . i_c . \Delta t \qquad \ldots\ldots(2)$$

If Δt were made sufficiently small, the resulting accuracy is still high. (The quantities C and i_c need not be linear with v.)

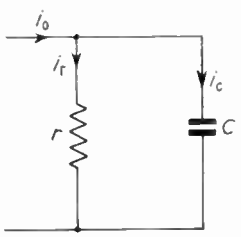

Fig. 2. Circuit used in description of template methods.

For the solution, a template is constructed in the form of a right-angle triangle, with an angle θ equal to arc tan $(1/\Delta t)$, and with multiplying constants suitably chosen to turn the angle into a convenient value. For example, applying equation (2) to the circuit of Fig. 2, and with a suitable value for angle θ, the solution is as shown in Fig. 3, where points 0, 1, 2, ..., n, denote the value of v at time intervals 0, Δt, $2(\Delta t)$, ..., $n(\Delta t)$, respectively. The resulting response, Fig. 4, is in the expected exponential form. Of course, the usefulness of the method is demonstrated better with more complicated problems, involving non-linear elements.

With some problems, it may be necessary to use two or more templates, and to expand the scales in certain regions, in order to preserve the same accuracy throughout. The use of templates was felt to be

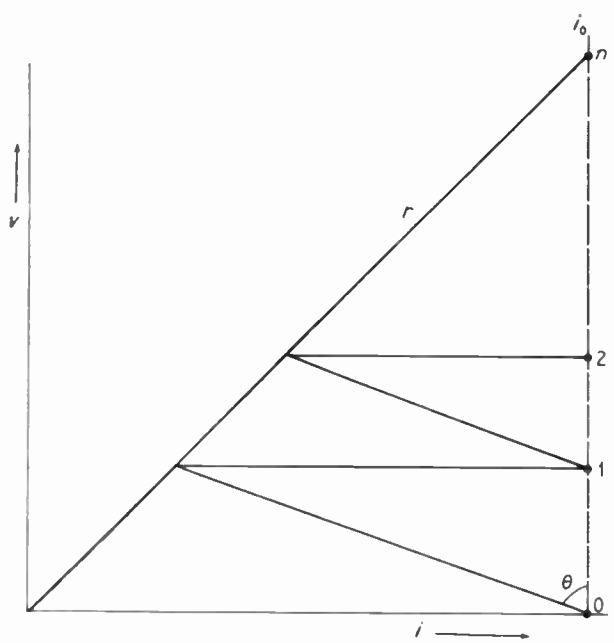

Fig. 3. Graphical derivation of response of circuit of Fig. 2.

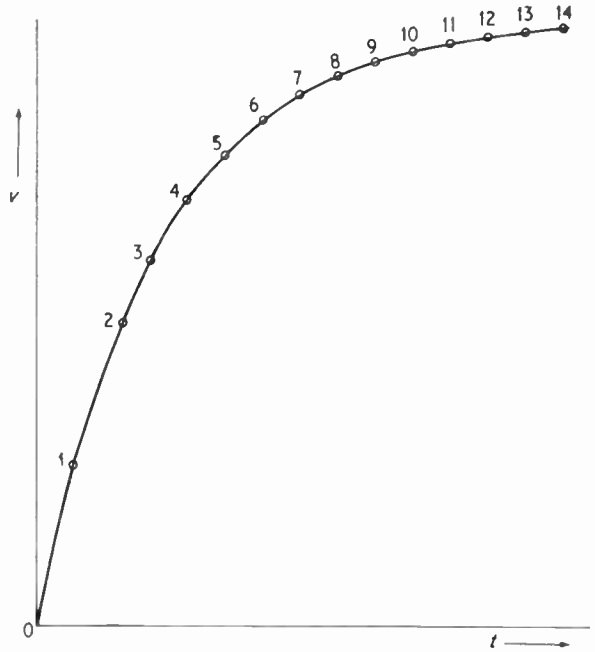

Fig. 4. Response of circuit of Fig. 2.

laborious, and an alternative method was devised, replacing this construction by a quicker slide rule calculation, equally accurate. That is, in the above example, the value of $\Delta t/C$ would be set on a slide rule, and for each point, the value of i_c determined from Fig. 3. The value of Δt is then conveniently read off the slide rule as

$$\frac{\Delta t}{C}(i_o - i_r).$$

While it may be argued that this method is limited by the usual accuracies of graphical solutions, this is offset to a degree by the absence of any piecewise linearization and/or the mathematical approximations of the analogue and digital computer solutions. It is also felt that graphical solutions are erroneously held by some to be of low accuracy. The opposite is more likely to be true, that the quickest graphical method offers comparable, if not superior, accuracy to that obtained from analytical solutions.

A further advantage of the graphical method, in the present application, is the ease with which the tunnel diode capacitance may be taken as varying, according to the law,

$$C_d = C_0(1 - v_d/\varepsilon)^{-\frac{1}{2}} \qquad \text{.....(3)}$$

where C_0 is the capacitance for zero applied voltage,
and ε is the energy gap,
rather than treated as a constant quantity.

3. Transistor Equivalent Circuit

The ASZ 21 transistor was chosen for the circuit, being comparable in performance and cost with the AEY 11 tunnel diode. Typical values used in the analysis are,

$$C_c = 3\,\text{pF},\ \omega_n = 3{\cdot}8 \times 10^9,\ \beta = 100$$

Since the switching times of the circuit are dependent more on the tunnel diode rather than on the transistor,[2,3] the high-frequency version of the T-equivalent circuit was chosen, being judged adequate for the purpose of analysis (Fig. 5). For convenience, the internal resistance of the source driving the circuit was added to $r_{bb'}$ of the transistor, forming r_{bt}.

The usual expression for the current gain, α, is:[7]

$$\alpha = \frac{\alpha_0}{1 + j\omega/\omega_n} \exp\left(-j\phi\frac{\omega}{\omega_n}\right) \qquad \text{.....(4)}$$

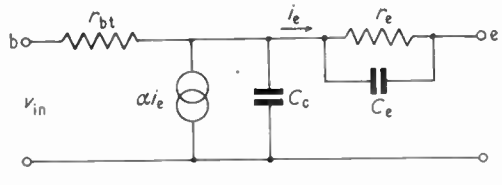

Fig. 5. High-frequency T-equivalent circuit.

The exponential part of eqn. (4) was found, for the cases considered, only to affect the initial transients of the switching waveforms. Since the overall switching times were of main interest, it was decided to exclude it from the above expression for α, leading to simpler expressions, though its effect could be taken into account through the use of poles and zeros in the form of a 'quad'.

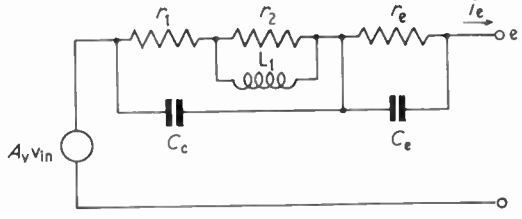

Fig. 6. New common-collector equivalent circuit.

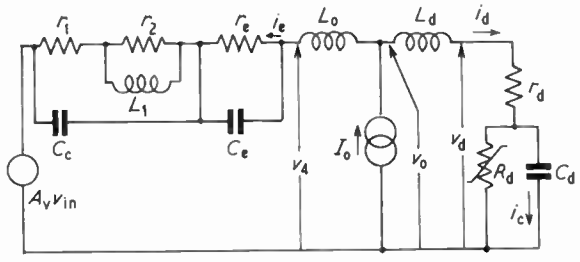

Fig. 7. Equivalent circuit of Fig. 1.

Applying Thevenin's theorem to the T-equivalent circuit of Fig. 5, yielded a new equivalent circuit for the common collector configuration (Fig. 6). The inductance L_1 in Fig. 6 simulates the rate of growth of current in the circuit, caused by the time constant of αi_e. The emitter resistance r_e was taken to be constant for simplicity, and because, during switching, its variation is negligible compared with the remaining resistance in the circuit. For optimum values of L_0, it will be shown that switching takes place at a fairly constant current, and therefore constant r_e. The complete equivalent circuit diagram for the circuit is shown in Fig. 7, where R_d represents the slope of the tunnel diode characteristic.

In order to maintain the comparison with previously derived trajectories,[3] the 4-section linearization of the tunnel diode characteristic was used (Fig. 8), and the diode capacitance C_d was assumed constant. It must be realized that the solution is equally simple using the practical tunnel diode characteristic, and with C_d varying.

For the solution, the following values were assumed, the units being ohms, nanohenrys, nanofarads, and

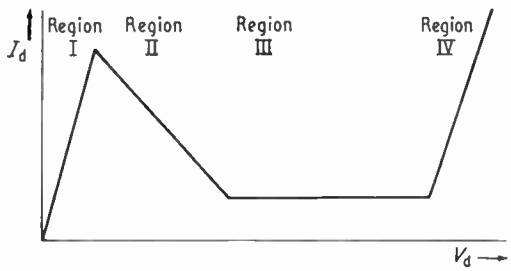

Fig. 8. Linearized tunnel diode characteristic.

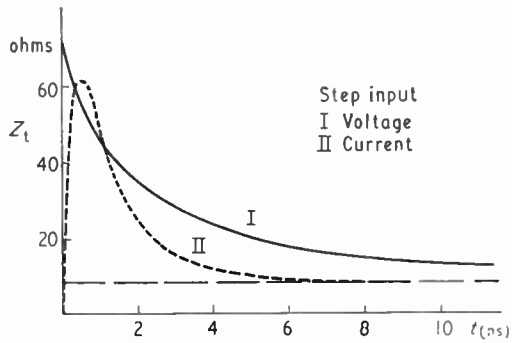

Fig. 9. Transistor response.

reciprocal nanoseconds (unless otherwise stated, these units have been used throughout):

$r_e = 5{\cdot}0$, $r_{bt} = 350$, $r_d = 1{\cdot}0$, $(1-\alpha_0) = 0{\cdot}01$,
$L_d = 10$, $L_0 = 10$ and 90, $C_d = 0{\cdot}030$, $C_c = 0{\cdot}0030$,
$R_d = 12{\cdot}5$, $-34{\cdot}0$, ∞, and $14{\cdot}0$ } for regions I–IV
$\omega_d = 2{\cdot}67$, $-0{\cdot}98$, $0{\cdot}00$, and $2{\cdot}38$ } respectively
$\omega_n = 3{\cdot}80$, $\omega_b = 0{\cdot}95$

Two values for L_0 were used, for a better estimation of the optimum value.

4. Qualitative Analysis of Circuit

Assuming a positive step input, A_v transforms this to an exponential input, with a time constant of $1/\omega_b$ (Fig. 7). The rate of current build-up is decided by the total circuit inductance L_t, which equals $L_0 + L_d$. Until this current build-up takes place, the value of the transistor output impedance Z_t is at or around zero, due to the uncharged collector and emitter capacitances, C_c and C_e respectively. But with the growth in current, these capacitances charge up and lead to a quick rise in the value of Z_t, up to its maximum level around 76 ohms. During this period, and because of the values of the circuit parameters, the transistor impedance Z_t absorbs a major part of the input voltage change, with the transistor output voltage staying nearly constant. In the meantime, the tunnel diode would have entered region II and started switching at a rate determined by C_d. With the tunnel diode switching faster than the change in the output voltage of the transistor, the voltage across L_t decreases, and stays around zero for a short period, during which the current remains constant. As the tunnel diode continues switching, the voltage across L_t increases in the opposite direction, with the transistor starting to load the tunnel diode. The graphical solution then centres on two equations.

$$\Delta i_d = v_1 . \Delta t / L_t \qquad \text{.....(5)}$$

and

$$\Delta v_d = i_e . \Delta t / C_d \qquad \text{.....(6)}$$

where v_1 is the voltage across L_t.

5. Variation of Transistor Output Impedance

In order to assess the nature of the loading of the transistor on the tunnel diode during switching, the behaviour of the transistor output impedance Z_t under different input conditions at the emitter should be analysed. This was done for ideal step inputs of voltage and current, Fig. 9, with zero and infinite source resistance respectively. As would be expected, the response of Z_t to a current-drive is faster than for a voltage-drive, with intermediate results for voltage inputs from sources of finite resistance.

6. Description of Graphical Method

In deriving the trajectory graphically, it is convenient to divide it into four sections or periods:

(a) The initial period, before the build-up in current.

(b) The period of current build-up, in which the transistor output impedance reaches its maximum value.

(c) The period during which the voltage across the inductance L_t is around zero, that is when the transistor and tunnel diode voltages increase at equivalent rates, with neither loading the other.

(d) The final period, when the transistor voltage changes very little in comparison with the tunnel diode voltage, constituting a loading on the tunnel diode by the transistor.

Items (a) and (b) depend on current growth, and should therefore be longer for $L_t = 110$ nH, than for $L_t = 20$ nH. (These values correspond to 100 nH and 10 nH for L_0, respectively). The most convenient way of evaluating the circuit response during these two periods was to use the approximation shown in Fig. 10. Since the voltage swing at the tunnel diode is now much smaller than that at the transistor output, the tunnel diode may be neglected, the inductance L_t forming the main load on the transistor. The increase in the current gain of the common collector is also small, due to the time-constant of γ, justifying the simplification to the transistor output impedance

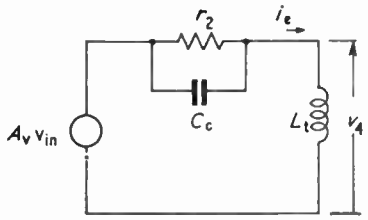

Fig. 10. Approximation to Fig. 7.

shown. Analysis of this simplified circuit, for L_t = 110 nH, leads to the following equations:

$$v_4 = v_{in}[5{\cdot}91 \exp(-0{\cdot}95t) - 5{\cdot}84 \exp(-0{\cdot}74t) - 0{\cdot}09 \exp(-4{\cdot}20t)] \quad \ldots\ldots(7)$$

and

$$i_e = \frac{v_{in}}{110}[1{\cdot}68 + 6{\cdot}22 \exp(-0{\cdot}95t) - 7{\cdot}90 \exp(-0{\cdot}74t)] \quad \ldots\ldots(8)$$

Equation (7) corresponds to an expression having two poles close to each other, with the third more distant. Thus v_4 rises quickly, due to the third pole, and then levels off for a period at $v_4 = 0{\cdot}50\, v_{in}$ due to the proximity of the other two poles. The expression for current, eqn. (8), does not include the third pole of eqn. (7), and so there is no rapid rise in current. Comparison of the above two equations shows that the poles which lead to the period of approximately constant voltage are those which cause a slow rise in current. One could then describe the two periods (a) and (b) respectively, as a period of quick voltage rise at nearly constant current, followed by a period of slow current rise at nearly constant voltage. A plot of these results for v_{in} = 0·4 volts, is shown in Fig. 11, along with a possible approximation for each trajectory, which may be used for simplicity, with little error.

A further useful property of eqns. (7) and (8) is their normalization in v_{in}. This means that the shape of the v_4 trajectory for periods (a) and (b) should be the same for any input voltage. Reference to analogue computer results[3] confirms this statement. This property serves as a 'time-saver' when trajectories for different amplitudes of input voltage are required.

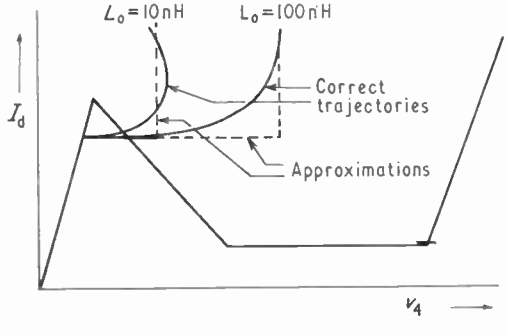

Fig. 11. Initial part of v_4 trajectories.

Analysing the case for L_t = 20 nH, gives

$$v_4 = v_{in}[0{\cdot}293\{\exp(-0{\cdot}95t) - (\cos 3{\cdot}31t - 0{\cdot}59 \sin 3{\cdot}31t)\exp(-2{\cdot}38t)\}] \quad \ldots\ldots(9)$$

This yields similar results to that for L_t = 110 nH, except that:

(1) during period (b) the voltage v_4 stabilizes around $0{\cdot}20\, v_{in}$.

(2) during this period, voltage v_4 has some ripple, caused by the lower circuit inductance bringing in the sinusoidal terms of eqn. (9).

The above analysis shows good agreement with the results obtained with the analogue computer.[3]

It has been shown that during period (b), the voltage available for increasing the current i_e in the circuit is about $0{\cdot}5\, v_{in}$ and $0{\cdot}2\, v_{in}$ for the L_t = 110 and L_t = 20 nH, respectively. That is, the voltage for the larger inductance is 2·5 times higher, so that even though the ratio of the two inductances is 5·5, the ratio of the delays in periods (a) and (b), for the two inductances, is just over 2, or (5·5/2·5). This again agrees with the analogue computer results. Considering the other advantages of the higher inductance, this extra delay can be tolerated.

Period (b) lasts until the trajectories for the voltages v_4 and v_0 merge into each other, when the voltage across the inductance L_t is approximately zero. This corresponds to the start of period (c), the duration of which depends on the relative rate of change of the transistor output voltage and the switching tunnel diode voltage. The larger the input voltage, the faster is the switching, and the longer is period (c). During this relatively short period, the transistor and tunnel diode do not load each other, so that the effective transistor output impedance Z_t remains constant near its maximum value.

Once the tunnel diode voltage switches faster than the change in the output voltage of the transistor, the latter begins to load the tunnel diode, and period (d) begins. The effective impedance Z_t now slowly decreases to its d.c. value, r_{ee}. For the conditions shown in Fig. 9, this decrease in Z_t takes from 5 to 20 ns, and the present case is nearer the 20 ns figure. This interval is large compared with the remaining period of any possible trajectory, making it possible to assume, without much error, an average value for Z_t.

This completes the construction of the v_4 trajectory of Fig. 12, where the linear approximations of Fig. 11 are conveniently used for sections (i) and (ii). The

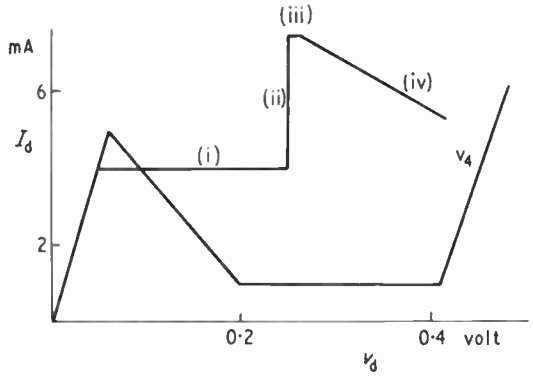

Fig. 12. v_4 trajectory.

quantities v_1 and i_c of eqns. (5) and (6) are now known, and from them the v_0 trajectory can be found.

7. Example of Graphical Solution

The solution for the added inductance $L_0 = 100$ nH, and a 0·3 volts step input will be outlined below (Fig. 13).

The time interval Δt was fixed at 0·5 ns, this being a reasonable figure for both speed and accuracy. Near the boundary regions, smaller intervals were used. With $L_t = 110$ nH and $C_d = 0{\cdot}03$ nF, eqns. (5) and (6) simplify to

$$\Delta i_d = v_1/220 \qquad \text{.....(5a)}$$

and

$$\Delta v_d = 16{\cdot}6 i_c \qquad \text{.....(6a)}$$

with v_1 and i_c representing average values during the time interval taken.

The above equations strictly yield the v_d trajectory. To obtain the v_0 trajectory, one must add $(v_4 - v_d)L_d/(L_0 + L_d)$ to the v_d trajectory. This addition is also carried out graphically.

Equations (7) and (8) indicate period (a) to be approximately equal to 0·7 ns. During this interval, the input voltage $A_v v_{in}$ reaches about half its final value which in this case is 146 mV. Therefore, the length of part (i) of the v_4 trajectory is made equal to 146 mV, and from its extremity, part (ii) is drawn corresponding to period (b). This determines v_1 of eqn. (5a) for the first part of the v_d trajectory, and the graphical solution for this trajectory can proceed. Once this trajectory enters region II (Fig. 8) the current i_c becomes appreciable, and eqn. (6a) is applied as well as eqn. (5a). Over the major part of any trajectory, one of these two equations becomes dominant, and should be used first in the evaluation. But there are also regions of a trajectory where both equations are of equal importance, and so affect each other. In such instances, more than one trial may be necessary before both equations may be satisfied simultaneously.

In the present example, period (b) lasts until 2·80 ns approximately, that is, until the v_4 and v_d trajectories converge. This necessarily limits the duration of period (c) to under 0·2 ns, since after 2·8 ns, the effective input voltage $A_v v_{in}$, with time constant $1/\omega_b = 1{\cdot}05$ ns, changes very little.

Period (d) follows, during which $A_v v_{in}$ is constant, and the impedance Z_t is also taken constant at an average value of about 70 ohms (see Section 6). Establishing the v_4 trajectory in this region ensures the completion of the v_0 trajectory.

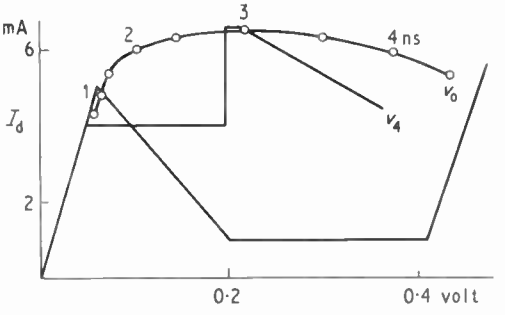

Fig. 13. Derivation of v_0 trajectory.

Similar graphical techniques led to the other solutions of Fig. 14. For $L_0 = 100$ nH, periods (a) and (b) are similar for the different input voltages, as was shown in Section 6. The same holds for the 10 nH trajectories, except that v_4 stabilizes around $0{\cdot}2\, v_{in}$ in period (b). It is seen that, considering the speed of carrying out these graphical solutions, the accuracy of the method is high, as a comparison with the analogue results of Fig. 15 would indicate.

8. Errors in the Graphical Method

The main error in the trajectories of Fig. 14, upon comparing them with the analogue results of Fig. 15 is shown to be in period (d), and for low input voltages. This is due, as expected, to the assumption of a constant value for transistor impedance Z_t. And since the period (d) is longest for the lowest input, the error due to this assumption becomes more noticeable for this value. For quick evaluations, this small error can be easily tolerated, and more so if the voltage response, and not the complete trajectory, is only considered. This is justifiable, as the output of the circuit is the voltage v_o.

But for an improvement in the accuracy of region (d) for small inputs, a more detailed study of the behaviour of the impedance Z_t under different conditions than those of Section 5 was made, to include

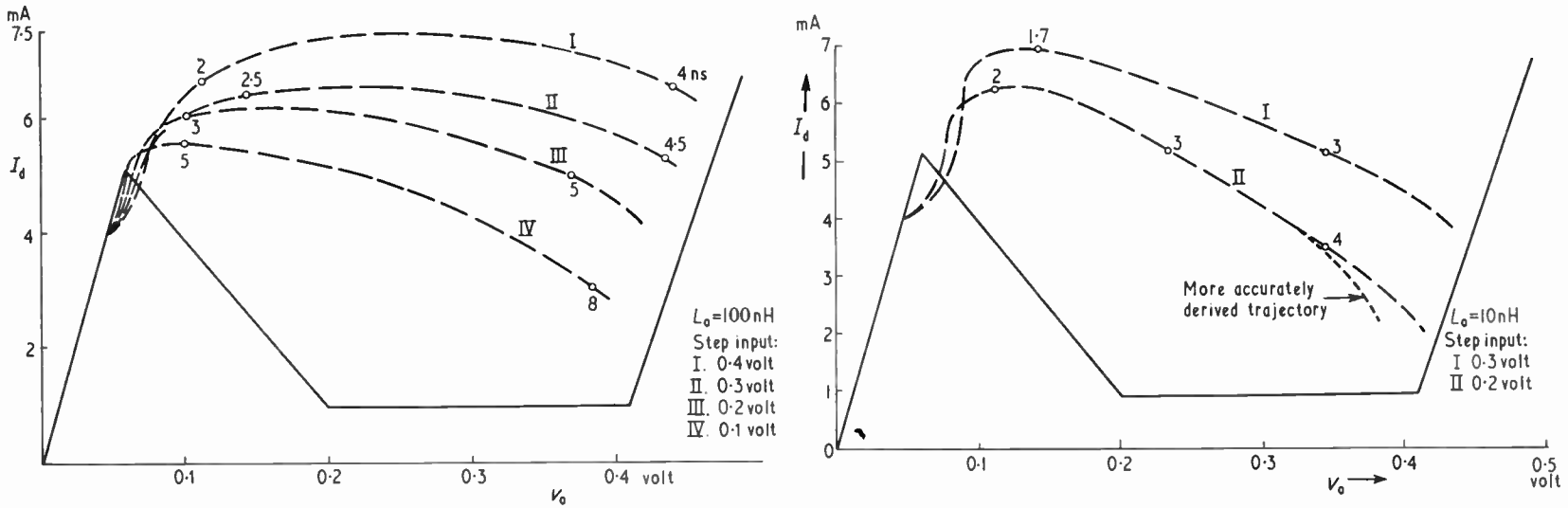

Fig. 14. Graphically-derived trajectories.

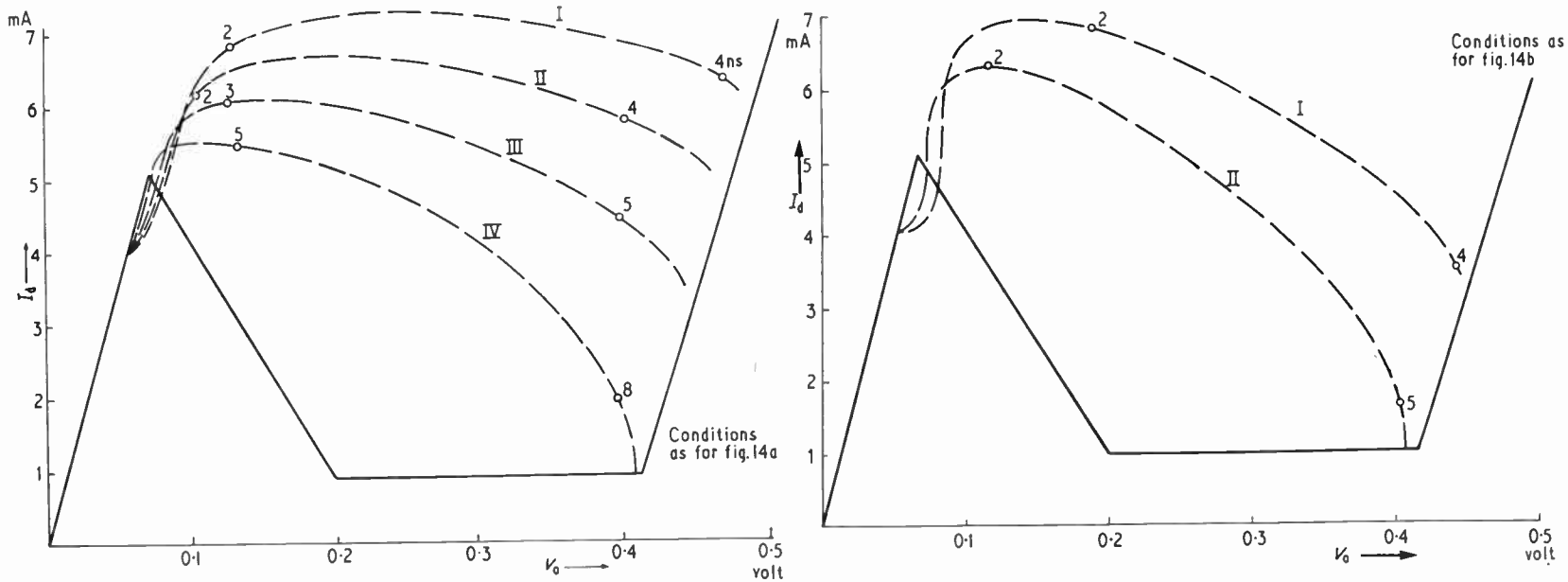

Fig. 15. Trajectories derived using analogue computer.

the finite rise-time of the input and the different possible source impedances. These results gave an average variation in Z_t from about 70 to 40 ohms in 2 ns, and reaching the d.c. value of r_{ee}, 8·5 ohms, in 12 to 15 ns.

Using these results, part (d) of the 0·2 volt trajectory of Fig. 14(b) was reconstructed, and is plotted in the figure. It is seen that better agreement with the analogue results has been achieved.

9. Conclusion

It is obvious that the accuracy of the method improves with smaller time intervals, but the 0·5 ns interval chosen for the described solutions was found to be a good compromise between speed and accuracy. This interval was varied, where necessary, to preserve the same accuracy. The aim behind this graphical method was not to obtain a rigorous solution, and the assumptions made in this particular example may have to be varied to suit other values of circuit parameters. A more detailed study could also develop the method further, improving the all-round accuracy, but this was not the object of the exercise. Yet enough has been described to show the versatility of the graphical approach to this and other similar problems.

10. Acknowledgment

This paper is based on part of a section of a thesis submitted for the Ph.D. degree at the Queen's University of Belfast. The author would like to record his appreciation of the advice and encouragement of his supervisor, Dr. W. D. Ryan.

11. References

1. G. W. Neff, S. A. Butler and D. L. Critchlow, 'Esaki diode logic circuits', *Trans. Inst. Radio Engrs on Electronic Computers*, EC-9, pp. 423–429, December 1960.
2. H. Madani and W. D. Ryan, 'Transistor–tunnel diode ternary logic circuits', (to be published).
3. H. Madani, W. D. Ryan and R. E. King, 'Transistor–tunnel diode circuit response by direct analogue', (to be published).
4. P. S. Hsia, 'A graphical analysis for non-linear systems', *Proc. Instn. Elect. Engng*, Part II, **99**, pp. 125–131, 1952.
5. S. B. Geller and P. A. Mantek, 'Tunnel diode large-signal equivalent circuit study and the solutions of its nonlinear differential equations', *J. Res. Nat. Bur. Stands*, (U.S.A.) **66D**, No. 1, pp. 45–50, January–March 1962.
6. A. Preisman, 'Graphical Constructions for Vacuum-Tube Circuits', (McGraw-Hill, New York, 1943).
7. J. te Winkel, 'Drift transistors', *Electronic and Radio Engineer*, **36**, p. 280, August 1959.

Manuscript first received by the Institution on 6th August 1965 and in revised form on 19th April 1966. (Paper No. 1074/C88).

(*continued from facing page*)

'Microwave and Optical Generation and Amplification'—12th to 16th September

The International Conference on Microwave Tubes has been held in several centres throughout the World, but for the Sixth Conference, the first to be held in Great Britain, the title was widened to take account of the great extension of work in this field. The Conference was sponsored by the Electronics Division of the Institution of Electrical Engineers and the I.E.R.E. and was held in three large lecture theatres of the University of Cambridge.

The extremely broad subject was divided into a number of groups, notably Quantum Electronic Devices, Crossed Field Devices, Linear Beam Devices, Solid State Devices and Gaseous Plasma Devices. Over 100 papers were dealt with in the course of five days which were attended by over 450 scientists and engineers—half of them from overseas.

Despite the growing range of applications for solid state devices, as was indicated at the 1965 I.E.R.E.-I.E.E. Conference on 'Microwave Applications of Semiconductors', the electron tube in its many different forms still provides a versatile and flexible means of generating microwaves at high power.

The Conference publication will be issued by the Institution of Electrical Engineers towards the end of this year and will cost £6 to members and £8 to others.

'Electrical Networks'—13th to 15th September

The University of Newcastle-upon-Tyne Electrical Engineering Department joined with the I.E.E. and I.E.R.E. and the Ministry of Technology in sponsoring a three-day Conference on Electrical Networks. This highly specialized field proved to be extremely well supported and some 200 persons attended, many of them from overseas. Papers were drawn from university and industrial laboratories in Great Britain and from Germany and the U.S.A. The Conference publication will be published by the Institution of Electrical Engineers later this year, price £2 to members of the sponsoring organizations and £3 to non-members (provisionally).

Institution Summer Conferences, 1966

During the past year an unusually large number of members of this Institution and of the Institution of Electrical Engineers have been engaged, either as authors or assisting Organizing Committees, in preparation for four large scale Conferences which took place in July and September. It is some indication of the wide ranging extent of electronics today that, although three of these Conferences took place in the same week in September, apparently very few engineers had professional interests which called for a choice to be made between attending one or other of these Conferences. The following reports highlight some of the features of the conferences.

'Applications of Thin Films in Electronic Engineering'—11th to 14th July

The first of the Joint I.E.R.E.-I.E.E. Conferences took place in London, at Imperial College of Science and Technology, South Kensington, and dealt most comprehensively with developments in one of the very important groups of techniques which will be employed in shaping the microelectronics circuits of the future (and indeed are already being employed in an extensive range of applications). Forty-three papers were presented and the sessions were grouped under the broad headings of 'The Preparation of Thin Films', 'Thin Film Elements and Integrated Circuits', 'Magnetic Films', 'Cryoelectric Films', and 'General Applications'; one of the most popular sessions was an informal discussion on 'Why Thin Film Circuits?' at which most of the 425 who had registered to attend were present, nearly 100 in an overflow theatre linked to the main lecture theatre by closed circuit television. Cost, availability, reliability, particularly in comparison with silicon integrated circuits, all received consideration, and the importance of 'thick' films was also mentioned.

The papers presented at the Conference were given in abstract form in May issue of *The Radio and Electronic Engineer* and it is expected that the final Proceedings of the Conference will be ready towards the end of the year. (The book containing the 27 papers preprinted may still be obtained from the I.E.R.E. at a cost of £6 which includes the supplement volume containing the remaining papers and reports of the discussions.)

'Electronic Engineering in Oceanography'—12th to 15th September

The key-note of the I.E.R.E. Conference which was held at Southampton University was summed up by the Address given by Sir Frederick Brundrett, K.C.B., K.B.E., Chairman of the Research and Development Committee of the White Fish Authority, in which he stressed the important part which electronic techniques had to play in promoting great efficiency in obtaining the extra food which the growing population of the world was making necessary. This was dealt with in many of the papers, while others discussed applications which could be regarded as more concerned with oceanography as 'science that is done at sea' or as a tool for the geologist wishing to find out more about the sea bed and what lies immediately beneath it.

A pleasing feature of this Conference was the high overseas participation: out of the 44 papers which were presented 17 were by authors from five countries outside the British Isles and the total registered attendance of 240 was divided in the ratio of 180 British Isles to 60 Overseas. In a Conference which aimed at bringing together workers from two distinct disciplines, the opportunities for informal discussion which its residential character provided were even more valuable than in a 'one discipline' Conference. Oceanographers and electronic engineers will undoubtedly find themselves working far more closely together in the coming years in finding out more about that vast part of the Earth's surface, about which we know little more than we do about space—Sir Frederick Brundrett's suggestion that oceanographic research is potentially a more rewarding field for Great Britain than space research will be regarded as the message of this Conference.

This note on the Oceanography Conference would not be complete without paying tribute to the great amount of assistance which the Institution received from the National Institute of Oceanography. In addition to contributing several papers and providing expertise for the Organizing Committee, a large party of delegates visited the Institute at the close of the Conference.

The outstanding social occasion of the Conference was the Dinner held in the Polygon Hotel, Southampton, when the President of the Institution, Colonel G. W. Raby, C.B.E., took the Chair. The toast to the Authors and Contributors to the Conference was proposed by Admiral of the Fleet Earl Mountbatten of Burma, K.G., in a characteristic speech in which he recalled the growing importance which scientific oceanography had assumed in the work of the Royal Navy. He also made reference to the collaboration between the National Electronics Research Council, of which he is the Chairman, and the White Fish Authority on work being carried out to increase productivity of the fishing fleets. The toast was replied to by Professor Dr. J. C. Schoenfeld of the Netherlands who spoke of the importance of such Conferences in co-ordinating the efforts of scientists and specialists in many different fields.

Summaries of the papers presented were published in the August issue of *The Radio and Electronic Engineer*, and as in the case of the Thin Films Conference publication, the preprint volume is available at a charge of £6 which will include the provision of the supplement volume containing the remainder of the papers and discussion reports.

(*continued on facing page*)

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.

COMPUTER SIMULATION OF A COMMUNICATION CHANNEL

A possible mathematical model of a random multi-path communication channel using the long-distance tropospheric propagation of ultrashort waves is described.

This model may be used to investigate the statistical properties of channel and signal, as well as the noise immunity and information capacity of tropospheric links, by employing Monte-Carlo techniques. A block diagram of an algorithm for computer simulation of the channel model is given, as well as practical forms obtained by modelling the frequency and time random characteristics of a channel.

The same mathematical model for a random multi-path channel may also be used for simulating long-distance stratospheric and ionospheric-scattering communication channels, channels with radio-wave scattering by an undulating sea surface, etc.

'Computer simulation of a troposcatter multi-path communication channel', I. P. Levshin and A. V. Prosin. *Telecommunications and Radio Engineering* (Translation of *Elektrosvyaz Radiotekhnika*), **21**, No. 2, pp. 64–70, February 1966.

A DIGITAL PHASE METER

The logical design principles for a system of phase angle measurement for use in conjunction with electronic aids to navigation or surveying have been described in an Australian paper.

The principle employed in measuring the phase difference between two continuous wave signals is the use of a gated reversible electronic counter to measure the time interval between corresponding points on the voltage waveform of the two input signals as a *fraction* of the time period of either signal. This technique obviates the necessity for any other fixed frequency relationship between the signals being measured and the counting pulse train used in the phase meter.

The counting pulses are gated into the reversible counter using a novel gating logic which enables the phase meter to distinguish between lead and lag angles and also to register, in the correct sense, changes of phase angle corresponding to complete cycles of 360°. This system of gating logic also provides for the smoothing of phase jitter in the input signal by averaging over a number of cycles. The smoothing process applies correctly even for 0° or 180° phase difference in contrast to some other systems which have been described.

'A digital phase meter for electronic navigational aids', A. Dunworth, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **27**, No. 8, pp. 213–221, August 1966.

DIODE MATRICES FOR ADDRESS DECODING

Digital data storage units with free access to the storage cell call for an address-decoding device for selecting the required address. This decoding is effected with AND-switching circuits that are preferably arranged in matrix form. Under the assumption that the encoded address is available in mixed encoding (simultaneous adoption of a 1-out of-u-code and 1-out of-v-code) the authors of a German paper show how the configuration of diode matrices that leads to the least amount of equipment can be determined by calculation or graphical methods. Single-stage decoding is discussed with multi-dimensional diode matrix, and multi-stage decoding with two-dimensional diode matrices, the latter in pyramid and tree form. From the decoding schemes thus developed the number of possible combinations and the layout for a minimum of necessary equipment can be determined. The amount of decoding facilities required for an address encoded in a pure binary and in a decadic form is then given.

It is also shown briefly how the number of components required at the address end can be reduced by combined word selection at the address and information ends.

Finally, the replacement of the diodes by other selecting elements which allow a linking of logic and power input coupling is pointed out and this can serve the purpose of directly driving a storage or coupling matrix.

'Determination of the requirements for the use of diode matrices for address decoding', O. Feustel, *Archiv der Elektrischen Übertragung*, **20**, No. 1, pp. 42–52, January 1966.

STORAGE OF RADAR INFORMATION

Increased application of traffic control radars in centralized surveillance and control of shipping on waterways has led to the suggestion for storing radar signals on magnetic tape for the purpose of documenting traffic events. Wideband recordings are very expensive and can hardly be taken into consideration.

A German paper describes an experimental system for compressing the bandwidth of radar signals from traffic control radar equipment. Without noticeable loss in picture quality the bandwidth of more than 5 MHz is reduced to about 200 kHz. Whilst analogue stores have so far been used in known methods for an intermediate storage of signals, in the new method the radar signal is digitized and recorded in a ferrite core store. After a conversion of the binary stored signal back into an analogue signal the latter can be recorded on magnetic tape or displayed on a display for bandwidth-compressed signals.

'Digital bandwidth compressor for radar signals', W. H. Schönfeld, H. Gillmann and H. E. Speckler, *Nachrichtentechnische Zeitschrift*, **19**, No. 6, pp. 313–321, June 1966.