JOURNAL OF

The British Institution of Radio Engineers

(FOUNDED IN 1925 – INCORPORATED IN 1932)

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

(from the objects of the Institution)

Vol. XII (New Series) No. 1 JANUARY 1952

NOTICES

Brit.I.R.E. Building Appeal

With the increase in size and activities of the Institution, the present headquarters are becoming inadequate. It has always been a handicap to the Institution that it does not own a building for housing offices, library, committee rooms and lecture theatre. The present headquarters building is rented and the London Section meetings also take place in a rented lecture theatre.

An appeal is now being launched by the Institution to the radio industry for funds which would enable the Institution to purchase its own building and in a later issue particulars of the appeal will be published.

Indian Sections

As announced in the September Journal, the visit of the General Secretary, Mr. G. D. Clifford, to India, has as its main purpose the formation of local sections of the Institution under the general ægis of an Indian Advisory Committee.

The visit has aroused great interest among all sections of the Indian radio industry and in government establishments, and Mr. Clifford has addressed meetings of members and others in many centres all over the country.

So far, local sections have been formed in Bombay, Delhi, Madras, Bangalore and Calcutta and regular meetings will be held in these centres for the reading of technical papers. Such progress is particularly welcome as circumstances have led to the postponement of the Convention which was to have taken place in Bombay in February.

Full details of the Local Sections and their activities will be given in a future issue of the Journal.

International Radio and Electronics Exhibition of India

The Radio and Electronics Exhibition which was originally to take place in Bombay between 9th and 29th February, 1952, has had to be postponed because of the failure of the monsoon. Drastic cuts in water and electricity supplies in the City of Bombay have been necessary, and it is likely that there will be a serious famine in the State.

The organizers of this exhibition have announced that it will be held from 10th to 30th November, 1952, and the hope is expressed that the delay may provide exhibitors with useful extra time to prepare their exhibits.

Appointments Register

The attention of all members, particularly those wishing to change their employment, is drawn to the Appointments Register maintained by the Institution. This Register, which is licensed by the L.C.C., was referred to in the 1951 edition of the Year Book.

As from the beginning of 1952, it is proposed to circulate regularly lists of vacancies to all members whose names are on the register and employers will similarly receive lists of situations wanted.

In order that the register should be of the maximum value, both to employers and employees, all those who wish to avail themselves of its services (which are entirely free) should write to the Appointments Registrar at the Institution.

Details furnished by employers should include full particulars of the posts to be filled, experience, age and qualifications required and the salary range (where applicable). Members should ask for a form " A."

E.R.A. Report

A report entitled "A Single Frequency Instrument for the Measurement of Interference with Television Reception due to Ignition Systems" has just been issued by the British Electrical and Allied Industries Research Association.

This report describes means by which a simple and readily producible instrument can be employed for assessing the interference with television reception caused by motor vehicle ignition systems. As a result, recommendations have been made to the B.S.I. for incorporation in the appropriate Code of Practice (C.P. 1001 : 1947), and suitable equipment has been commercially developed.

Copies of the report (Ref. M/T102) can be obtained from the E.R.A., Thorncroft Manor, Dorking Road, Leatherhead, Surrey.

Binding of Journals

Members who wish to have their 1951 Journals bound are asked to send them to the Librarian, 9 Bedford Square, London, W.C.1. A charge of 12s. 6d. is made for binding, which will include an index.

SOME NAVIGATIONAL AND AIR TRAFFIC CONTROL PROBLEMS OF CIVIL AVIATION AND THE APPLICATION OF RADIO TO THEIR REDUCTION*

by

G. W. Stallibrass†

A Paper presented at the Fourth Session of the 1951 *Radio Convention on July 27th at University College, Southampton.*

SUMMARY

This paper reviews certain Navigational and Air Traffic Control problems, and discusses the application of radio to them. The international background is mentioned, but details are left to other papers.

The problems are considered under the different phases of flight in which they occur, and possible developments to meet them are outlined. These include consideration of the use of long-distance and short-distance radar surveillance in present Air Traffic Control Systems, and to developments needed in equipments and the techniques of use. Reference is made to current thinking on these problems in the United States.

A general account is given of the requirements likely to arise from the introduction of turbine-engined aircraft into civil aviation.

The particular problems of long- and short-distance routes over sparsely inhabited territories are considered and some factors mentioned relating to general international standardization on particular facilities.

1. Introduction

This paper deals with certain problems now encountered or coming into view in the fields of Navigation and Air Traffic Control. In order to understand the trends, it is helpful to go back a little way and see what has led up to them. It should be added that the opinions expressed are not necessarily the official views of the Ministry of Civil Aviation and should not be read as such.

A good starting point is the report of the Special Radio Technical Division ("COT") of the International Civil Aviation Organization which held its only meeting, so far, in Montreal during 1946. The general outlines of ICAO thought at that time were given in a paper‡ delivered to this Institution by Captain Hunt, but some important factors have changed since then. The early post-war discussions of P.I.C.A.O. and, later, I.C.A.O. on these matters were characterized by hopes which have since proved too sanguine. The rapid developments of radio during the war led to a belief that certain technical problems would be solved more quickly than has proved possible; and there was a justifiable expectation that, after the war, full international co-operation and a rapid increase in the world's technical resources would soon make it technically and economically practicable to introduce greatly improved systems on a wide scale. It is now necessary to recognize the melancholy fact that the far-reaching results of increasing international tension, and the economic difficulties in which Europe and other parts of the world soon found themselves after the war, have been reflected partly in a slowing down of the rate of development of equipment, and partly in delays to the installation of new systems because of shortages of resources and technicians. The effect is that, in 1951, en route Navigational and Air Traffic Control procedures over most of the world are still based on radio facilities that were recognized as ripe for replacement in 1947. Further, the failure of certain new equipments to be developed to the standard originally envisaged, and clearer realization of the costs of providing numerous new installations, has introduced an element of uncertainty so that the absolute necessity of world-wide standardization of aids, at one time considered axiomatic, is now being re-examined. The airlines themselves now have doubts of the economic justification for installing certain airborne equipmen whose development has been strongly urged in the past.

* Manuscript received June 30th, 1951.
† Directorate of Navigational Services (Control and Navigation), Ministry of Civil Aviation.
U.D.C. No. 656.7:621.39.

‡ Capt. V. A. M. Hunt. "Functional Requirements for Radio Aids to Civil Aviation." *J.Brit.I.R.E.*, 8, March-April, 1948, p. 41.

It will be convenient to review the situation under the general headings of High Density Areas and Low Density Areas. The different considerations likely to affect operations in these areas may be illustrated by comparing conditions in the United Kingdom and, say, Nigeria. Nigeria has over four times the surface area of Great Britain and about half its population. The mileage of routes to and within it is only about 1/16th of the figure for Great Britain, and the ratio for the weekly scheduled miles flown shows an even greater discrepancy of over 40 to 1. As will be seen later, these factors are indicative that several of the problems encountered in two such areas will be very different and will require different solutions.

2. High Density Areas

In these areas a high proportion of services will be flown on short stages, there will be a density of traffic along a network of routes, and congestion may occur at the main aerodromes. Congestion is used here to indicate that the movements of scheduled aircraft are sufficiently numerous to require the adoption of special procedures, if delays and the risk of collision are to be kept to tolerable limits. Navigation is no longer simply a problem of flying direct from one aerodrome to another, but has become a matter of timing and routeing the flight, in accordance with a previously notified Flight Plan or the in-flight requirements of Air Traffic Control. In short, the problems of navigation along the main routes have become inseparable from the requirements of an effective Air Traffic Control system.

2.1. *En Route Navigation*

The general problems of this type of area may be illustrated in Europe, and in particular by our own experience in the United Kingdom. On several European routes there are now corridors of controlled air-space known as Airways, and aircraft desiring to fly within them under Instrument Flight Rules must do so in accordance with an air traffic clearance specifying route, timing, and altitudes. The radio aids to mark them are Medium-Frequency radio ranges and M.F. non-directional beacons used in conjunction with Automatic Direction Finders on the aircraft, although some aircraft still use Gee where cover is available. There are still some parts of Europe without airways, and others where their demarcation is by non-directional beacons only.

The Airways Scheme which now serves major routes in the United Kingdom is marked by M.F. ranges. The fact that a shortage of frequencies on the M.F. band seriously restricts the number and power of M.F. installations is probably well known; but the feature which is of particular note is that, even when skilfully used, ranges and beacons by themselves can only give an indication of track and not of position. If an indication of position is required along a leg of a radio range, it requires the installation of a non-directional beacon or a V.H.F. fan marker. There is obviously a limit to the number of marker beacons that it is practicable to install, and the use of non-directional beacons for this purpose is limited by the frequency difficulty just mentioned. This severely limits the position information obtainable from aircraft as they pass along the airways. So far as lateral tracking separation is concerned, beyond a distance of about 15 miles from the range station, it is practicable to establish two tracks in opposing directions by means of what is known as "right side separation"; but on the whole it is fair to say that, with Airways marked by radio ranges, Air Traffic Control cannot count on being able to give aircraft lateral separation wherever it is needed.

Within these limits, radio ranges can and do give very useful assistance. It is apparent, however, that their services cannot meet all requirements when two busy Airways join and it is impracticable to assign vertical separation to a succession of aircraft passing through the junction point: when fast aircraft are overtaking slower aircraft and, at some stage, descending through their altitude to reach a holding pattern ahead of them: and when aircraft with long ascent and descent paths, such as the later models of certain piston-engined types and probably all turbine-engined aircraft, require to go through the successive altitudes of an Airway steadily during a period of anything from 15 to 40 minutes. Earlier this year the accuracy of position reports was analyzed on certain U.K. airways, using radar as a check on the position reports transmitted by aircraft. It was found that most aircraft could not be relied on to be closer than within ± 3 nautical miles of their reported position along the airway, and within $2\frac{1}{2}$ nautical miles of the centre of the airway. The accuracy of reports based on non-directional beacons was greater than for reports based on fan markers, and there were a few gross errors of 7 to 8 nautical miles. So long as Air Traffic Control is dependent for its knowledge of the relative positions of aircraft on the times at which they last reported over a reporting point, or on the E.T.A.s given for their arrival at the next reporting point, the time intervals used for standard separations must be large enough to take account of inherent inaccuracies. In the absence of radar information, they may be as much as 10 to 15 minutes, which can lead to unacceptable delays. Variations of track-indicating aids, such as VAR—which is a combined installation giving aurally indicated "legs" in two directions and visually indicated "legs" in two other directions—and the V.H.F. Omnirange (VOR) as introduced in the United States, do not remove these difficulties.

The fundamental requirement is for an aid that will enable an aircraft to determine its position instantly and continuously, and to make good any track required within coverage of the aid concerned. This requirement was proposed by the British and accepted at an International Conference as long ago as April, 1946. It is not the whole requirement, for there remains the problem of providing the position and track information to Air Traffic Control. This, however, is dealt with later.

2.2. *Terminal Areas*

At present, congestion mostly occurs in Terminal Areas and is likely to remain a major problem there, since the constant effort is to speed up landing and departure rates. Before considering the sequence of flight for incoming aircraft in the Terminal Area, it is worth spending a little time on one feature that will be referred to frequently, namely the Holding Pattern. This is used for regulating the final arrival times of aircraft by passing them through a set of altitudes over a point marked by a radio aid, and adjusting their times of leaving each altitude in order to secure an even flow from the base of the Holding Pattern to the runway. A vertical type of Holding Pattern like this is known colloquially as a "stack." The opinion has sometimes been expressed that, if Air Traffic Control were really efficient, there would be no need for stacks at all. To many people the idea of the stack is associated only with delays at the end of a flight, during which the aircraft flies around with nothing to be seen from the windows and the passengers become increasingly impatient or bored. This view, however understandable, is not borne out by the facts. It must be recognized that the arrival times of aircraft have to be regulated at some stage during their flight. However good the navigational aids provided, and however good the capabilities of the pilots, aircraft bound for London, for example, take off from aerodromes situated in many different parts of Europe and overseas so that their departure times cannot be regulated solely by regard to arriving at London Airport in a steady flow at intervals of, say, three minutes. Furthermore, there are economic limitations to the variations of speed that these aircraft can achieve en route, and they are flying through a medium whose conditions cannot be forecast by any meteorological service to a continuously high shade of accuracy. Although good work has already been done in endeavouring to adjust arrival schedules to prevent disproportionate bunching at particular aerodromes (e.g. at Northolt), these schedules cannot be maintained to the fine shade of accuracy required for smooth self-adjusted traffic flow. Present facilities do not make it practicable to zigzag aircraft about en route, in order to delay their arrivals; and, even given first-class radar and navigational facilities to do this, there would be limitations in the extent to which this could be done. The alternative is to make final adjustments to arrival times by means of a stack fairly near to the aerodrome. This is the first effective method open to us at present and it is likely to remain in some form or another indefinitely. At its worst, the additional safety factor achieved and the overall landing rate obtained through a stack should be considerably greater than that obtainable through unadjusted "ad hoc" approaches in I.F.R. conditions. At its best, aircraft should be flowing downwards through a vertical stack with practically no delay at the successive levels.

The flight sequence for incoming aircraft may be divided into the following stages:—

(1) Initial Approach, from Airway to Holding Pattern.

(2) Holding Pattern.

(3) Intermediate Flight from Holding Pattern to "gate" (a point 6-8 miles out on the extended centre line of the runway).

(4) Final Approach, from "gate" to runway.

(5) Surface Movement, from runway to unloading bay.

2.2.1. Initial Approach

For the Initial approach from the Airway to the Holding Pattern, the objective is to get the incoming aircraft into the lowest vacant altitude in the stack. Since the flight from the last reporting point should be short, the accuracy of the E.T.A. in the stack is generally fairly good; but there is need for an effective and foolproof means of checking that the assigned altitude in the stack will be vacant, and of notifying this information to the Air Traffic Control Officer responsible for giving the aircraft clearance to the stack. This need will be referred to again later. Another problem is that, with existing navigational aids, aircraft departing from the aerodrome have to be routed over the same facilities as those used for marking the Holding Pattern. This means that the lowest altitudes over the radio facility concerned will be reserved for departing aircraft but, once past that facility, they will require to climb as rapidly as possible to their efficient operating altitude. This often requires passing through the altitudes of incoming aircraft and, even with a more flexible navigational system than radio ranges, radar surveillance of the airspace concerned is essential.

2.2.2. Holding Pattern

As mentioned earlier, the Holding Pattern itself is defined by radio aids. Typical patterns are of a racecourse shape based either on two non-directional radio beacons or on a radio range with a non-directional or fan marker beacon situated a few miles along one leg from the range station. On arrival in the Holding Pattern or before, pilots are given the estimated time at which they should expect to leave the Holding Pattern. This is based on the number of aircraft ahead of them, multiplied by the landing interval being aimed at, with such time allowances as are needed for take-offs from the destination aerodrome. The stack-leaving time may be amended subsequently to take account of the movement rate actually achieved, and the amended time will be passed to the aircraft before it reaches the "leaving" altitude. At the upper altitudes the "racecourse" pattern usually consists of straight sides of two minutes' flying with a rate one turn at each end, making a total of six minutes. At the lowest altitude and sometimes at the one above, the length of the sides is reduced to one minute, making a total of four minutes. Unfortunately the navigational aids do not enable aircraft to maintain their assigned stack-leaving times with sufficient accuracy, mainly because the pilots can only know exactly where they are twice during each pattern, i.e. when they are passing over one or other of the beacons concerned. The performance noted from radar observations taken so far may improve as aircrews obtain further experience on the system; but, at present, allowance has to be made for leaving errors to vary between one minute early and nearly 1½ minutes late. Many errors will be less than this, but, on occasion, they will be more.

2.2.3. Intermediate Flight from Holding Pattern to "Gate"

Various methods have been pursued to take out the effect of these variations and to achieve a smooth flow. Considerable trials were carried out at La Guardia Airport, New York, using an apparatus known as the G.R.S. Timer, into which wind speeds, aircraft speeds and final landing directions were fed by Control. Aircraft were passed to an "inner" stack near the aerodrome marked by two non-directional beacons and pilots were assigned stack-leaving times as notified previously. An important feature was that all aircraft were fitted with twin ADF. When the first aircraft reported leaving the stack, the apparatus was set in motion by a switch and, when the second aircraft reported over the stack-leaving point, actuation of another switch caused the Timer to indicate to the Control Officer the heading on which the aircraft should fly and the length of time for which it should fly before turning on to the final approach marked by an ILS Localizer, in order to secure the desired separation. The Control Officer would tell the aircraft the heading on which to fly when it passed the last holding beacon and, when a flashing red light operated in the Timer, would tell it to begin its turn on to final.

The G.R.S. Timer was withdrawn from use a few months ago, reputedly because it contributed delay to the operation of Control rather than reducing it. Presumably this means that, when traffic was dense and the Controller was most busy, he had no time to feed in the wind speeds and airspeeds and operate the equipment. The Americans have recently started to use surveillance radar at New York, which is of interest since we ourselves began using surveillance radar for controlling aircraft from the stack some two years ago. The landing systems in use at London and Northolt are designed to take advantage of the fact that, with radar, Control can see when an aircraft leaves a stack and can assign it one of a number of flight paths from stack to "gate" whose variations in length will reduce the effect on runway arrival of errors in stack-leaving times. These tracks are pre-computed to give the desired variations and the aircraft is assigned the one appropriate to the particular variation required. They can also be used to reduce the effects of wide variations in approach speeds, but cannot eliminate them entirely since the approach speeds of different aircraft types vary by as much as 60 knots in extreme instances.

A remaining difficulty is that when an aircraft has been given a selected track to fly from stack to gate, there are still appreciable variations between the actual and estimated times taken by aircraft to cover even this short distance. Over a period of about 6½ minutes' flying, the variations may be as much as a minute. At the moment there is no available computer that will enable the Controller to give changes of heading during the intermediate stage in order to meet these final variations. This problem is one of a number now receiving attention.

2.2.4. Final Approach

On the Final Approach phase, the Air Traffic Control problem arises from the fact that operators of large aircraft prefer to have a straight-in Final Approach of not less than six nautical miles, in order to give the pilots time to settle down on their various instruments, including ILS Localizer and Glide Path receivers, and to complete their preliminary landing drill. At the same time, even the Final Approach speeds of various aircraft types may differ by as much as 45 knots, and with a 30-knot headwind this can make a difference of about 1½ minutes on Final Approach alone. During this stage, it is impracticable for Control to take any adjusting action other than instructing one of the aircraft to overshoot. The navigational problem, as distinct from the Control one, is to provide an aid on which it is practicable for the pilot to approach the runway confidently and accurately down to a low altitude. Present practice has shown that, used in conjunction with good approach lighting systems, either well-operated GCA or ILS equipment is reasonably adequate for this and the demand on the pilot can be very greatly reduced by instruments of the Zero Reader type. These present ILS-derived information in terms of rate indications, thus making it much easier to approach and hold the glide path and localizer "beams."

2.2.5. Surface Movement

The size of large aerodromes makes it impracticable to control the flow of aircraft between the runways and dispersal and loading points on the aerodrome by visual means from a central point. Instances are already on record where aircraft have managed to approach the runway in fog and to land on it, but then have been unable to see sufficiently to taxi off it. There are many problems here in aerodrome and taxi-way designs but, so far as radio is concerned, it is clear that some better means than reporting positions and obtaining taxi clearances by R/T is concerned. The latter would be too slow, too cumbersome, and, at the busiest periods, probably impracticable. In the United Kingdom a system has been developed known as the Ground Traffic Indicator which is being tried out at London Airport. Actuated by magnetic detector coils in the taxi-ways and runways, it provides the Surface Movement Controller with information on a "mimic" illuminated display of the aerodrome system and enables him to clear aircraft on the block signalling system. His instructions are passed to the aircraft by illuminated traffic signals sited alongside strategic points on the taxi-tracks. An alternative line of approach, also being pursued in the U.K. and U.S.A., is to apply radar surveillance to this problem. Experiments are being carried out this summer at London Airport, using a Marine Radar Equipment operating on three centimetres. Final consideration of the relative merits of the two systems cannot be made until investigations have proceeded further. It may be that the ultimate choice will vary from aerodrome to aerodrome, depending on such matters as the importance and density of the air traffic concerned, the layout of the aerodrome in relation to Permanent Echoes, and problems of radar identification.

An equipment is needed for guiding aircraft with sufficient accuracy for them to find and follow taxi-ways rapidly, and with ease in very low visibility. Not much can be said on this point here, except that experiments have been conducted for several years with leader cables buried below the surface of a taxi-track or runway, and that no means has yet been produced that would be acceptable and practicable for civil aviation.

One further factor should be mentioned, common to all airborne operations. This is the requirement for reducing the risk of interference between military aircraft and civil aircraft. In a small country like the United Kingdom it is not practicable to achieve the degree of segregation of military and civil activities obtaining in a continent such as the United States. Furthermore, fighter activities and exercises must often take place near the densely populated localities that would have to be defended in time of war, and these are also the focal points for civil air services. This accentuates the need for radar surveillance and identification of aircraft referred to later.

So far this paper has dwelt more on problems arising from the present situation and the limitations of existing equipment. It would be a mistake to infer from this either that the existing facilities are bad in themselves or that the overall situation is very unsatisfactory. In fact, although much more needs to be done, particularly in certain parts of Europe, the Airways System recently introduced in the U.K. works very much better than anything available here before, and landing rates in bad visibility at the more important aerodromes are increasing steadily. However, the development of the air transport industry continually sets the pace for the development of radio facilities serving it; and if emphasis is placed on the limitations of existing equipment rather than on the very considerable service it provides, it is on the assumption that more interest will be centred on requirements for new equipment still to be produced. These can now be considered in more detail.

2.3. *En Route Problems* (*Medium and Short Range*)

It has already been stated that the main problems of Air Traffic Control, as directly reflected in the requirements for navigational aids, occur in areas where there is a preponderance of traffic operating on short or medium stages, although a proportion of the traffic may also come from long-haul operations. The first basic requirement is for a navigational system that will provide aircraft with continuous information of their position, presented in such a form that they can make good any track laid down without the need for careful calculation, and without requiring the presence of a specialist navigator among the crew. The system should also provide aircraft with continuous indications of their ground speed since this, in conjunction with the track flown, is essential if they are to maintain E.T.A.s accurately. One obvious way of doing this is to present the information pictorially—for example, by means of a continuous trace on a map with time markings injected into the display so that ground speed can be determined as easily and often as is required. This should also enable position reports to be made more promptly, since pilots can see the rate at which they are approaching a reporting point and can be ready to make their reports at the right moment. The equipment should be easily set up, should give a quick and clear indication of malfunctioning, and should not require frequent resetting or retuning in flight. It should be as immune as practicable from serious atmospheric interference—which is a point in favour of an aid on a Very High Frequency. On the other hand, it should provide cover not only to very high flying aircraft but also at the low altitudes likely to be followed in future by helicopters—which is a point in favour of something on a Low or Medium Frequency, since low cover on V.H.F. would require very numerous installations. The information it provides should be sufficiently accurate for aircraft to be spaced safely and reliably with a lateral separation of two or three miles, and this accuracy should not be achieved at the expense of having an excessive number of ground installations. The system should also be sufficiently flexible for new routes to be operated within its cover, or for aircraft to be routed round storms and temporary danger areas, without the need for changing the number or sites of ground installations. It is clearly advantageous if the system can serve shipping as well as aviation, as was done formerly with a limited number of M.F. D/F stations and non-directional beacons.

Finally, any new system installed on the scale required should be capable of meeting all foreseeable needs likely to arrive over the next 15 years or so. Once installed, it will be impossible for economic and practical reasons to replace it a few years later; and great care has to be taken to check that a system, which may be hailed at the moment as a great advance over existing aids, does not contain within it limitations that would choke the progress of Air Transport when they became manifest in a few years time.

The two most likely contenders for selection as the new navigational system are the V.H.F. Omni-Directional Range (VOR) on 112-118 Mc/s together with Distance Measuring Equipment (DME) on 1,000 Mc/s; and the Decca System of hyperbolic navigation, incorporating the Automatic Flight Log and operating on low frequencies (100 kc/s). A detailed appraisal of these systems would run into many pages and only a summary can be given here. The VOR displays continuous information of the position line between an aircraft and the beacon being received, and enables this to be shown either in the form of a bearing or with left/right indication for tracking to or from the beacon, on a selected bearing. It is an attractive instrument to fly on, and its attractiveness arises in part from its superiority over the old A.D.F. in presentation of information. Judging from extensive trials in America, the order of accuracy of the information is around ± 5 deg., but might be a good deal worse with poor aircraft receivers. The limitations on siting ground beacons are very stringent and for physical or economic reasons it cannot always be sited where it is wanted.

DME has been visualized as the partner of VOR in providing a full navigational system. It has been used successfully on 200 Mc/s (Rebecca/Eureka) for many years, but insufficient information is available yet to speak of its performance with meter presentation on 1,000 Mc/s, although it looks like being very attractive. Whereas the American aircraft sets will only give distance information, the British equipment will provide heading indication as well. Airborne DME equipment, made to the I.C.A.O. requirement, looks like costing and weighing a good deal, and there are now doubts in America as to whether the operators will equip their aircraft with DME for a long time. In this event the concept of a Range/Bearing (R-φ) system based on VOR/DME, previously urged by America for international adoption, would have to be revised.

The Decca system, which now includes the Automatic Flight Log, provides aircraft with a continuous record of their present position and past track by means of a trace drawn on a map. Over most of the areas in question the position accuracy will be better than a nautical mile, and will, in fact, provide better cloud-breaking facilities than any available to-day, apart from Runway Approach systems. Unlike VOR/DME, it provides cover down to low altitudes. From the point of view of long-term investment, this is essential because it is expected that, within the next 10 years, helicopter services will have developed considerably.

VOR has been recommended by I.C.A.O. as a standard aid "in localities where conditions of traffic density and low visibility necessitate a short distance aid to navigation for the efficient exercise of Air Traffic Control." The United Kingdom dissented from this recommendation as it doubted the adequacy of VOR for this purpose, and these doubts have since been confirmed by the inaccuracies disclosed by extensive trials in the United States and by examination of the costs of providing VOR/DME on the scale required in Europe. For these reasons, at the recent meeting of the I.C.A.O. COM Division, the U.K. took a leading part in discussing the advances that have recently been made in the development and application of Decca, and this resulted in a recommendation that the system should be further examined and evaluated operationally, preferably in the European region when so much cover already existed.

Broadly speaking, the advantages offered by Decca are greater accuracy and cover, pictorial presentation, about two-thirds of the cost to authorities supplying ground installations, and availability for shipping purposes. On the airborne side, it may be about half the cost and a little over half the weight of corresponding equipment for VOR/DME.

It is worth considering for the moment the environment in which such a navigational aid would be used. On the civil side, it is expected that there will be a network of helicopter services operated at 1,000-4,000 ft. They will probably be frequent. From about 3,000 ft. to 10,000 ft. there will be a considerable density of piston-engined aircraft; and from 10,000 ft. to 20,000 ft. there will be a minority of piston-engined aircraft and an increasing proportion of turbine-propeller aircraft. Finally, there will be a number of turbo jet aircraft operating at anything between 25,000 ft. and 40,000 ft.

A prominent feature will be the long ascents and descents of the turbine-engined aircraft. When B.E.A.C. operated the Viscount on experimental services to Paris and Edinburgh last year, the aircraft on the Paris run, with a flight time of less than 70 minutes and an operating altitude of under 15,000 ft., was only in level flight for about a fifth of the time. On the longer route to Edinburgh, with an operating altitude of about 18,000 ft., and a flight time of about 90 minutes, it was in level flight for something under half the time. The problem of changing altitudes is a difficult one for Control, since an aircraft can only be cleared quickly through the altitude of another aircraft on the same track if there is sufficiently recent and accurate information on the positions of the aircraft concerned for the intersection risk to be taken out in the Flight clearance. It becomes essential to provide lateral separation of aircraft earlier than was foreseen several years ago, when it was thought that lateral separation would only be required when the densities of movements were very high indeed.

On a smaller scale this problem is already present on some portions of routes used by aircraft with such contrasting performances as the Dakota and the Convair 240. The first need is for the navigational aid to enable aircraft either to be given separate tracks for a good deal of the flight or to be put on to a separate track while they are changing altitude, after which they can come back to the "centre line" track again. This, however, only solves a part of the problem. At some point aircraft bound for the same destination have to be re-concentrated and, even with a good navigational aid, the constant position reporting needed for this and for dealing with aircraft crossing or joining the airways throws a heavy load on the R/T communications. This leads naturally to a consideration of the role of surveillance radar.

2.4. *Surveillance Radar*

Already, the use of long-distance surveillance radar enables the separation between aircraft to be reduced to lower figures than those based on Flight Plans and position reports only. By providing plan position information, radar enables Control to observe the movements of aircraft accurately and directly, and to give clearances based on fairly precise knowledge of aircraft position. It can also be used for initiating avoiding action if an "intruder" aircraft suddenly appears, for vectoring aircraft round storms, and for giving guidance to aircraft whose navigational equipment has failed until they are in the cover of GCA.

The equipment in use at the London Air Traffic Control Radar Unit ("London Radar") is Microwave Early Warning radar operating on a frequency of 10 cm. It is giving good service, and it seems almost ungrateful to point out the defects inherent in present types of radar. Nevertheless, they must be considered in order to see what improvements are required in future. The first limitation is that the basic display on present types of radar—a Plan Position Indicator—displays the positions of the aircraft as blips, without identification. Identification is a comparatively laborious process. A blip may be generally identified by the extent to which its position coincides with the estimated position of an aircraft for which a Flight Plan has been previously notified to Air Traffic Control. Additional information may be obtained by requesting the aircraft to transmit and then taking a bearing on it. Even this is not a positive indication if more than one blip is displayed in the area concerned, since the accuracy of ground D/F is less than that of the radar and is insufficient to discriminate between two aircraft close together. The final test is to ask the aircraft to make a turn for identification (usually about 60 deg.) and then to observe the behaviour of the blip on the P.P.I. Once the aircraft is identified, the Radar Controllers have to carry the identifications of the blips in their heads or mark the P.P.I. tube with a chinagraph pencil.

The long-term requirement is for aircraft to carry radar transponders that transmit their identity in code, either continuously or when interrogated. In practice, this is likely to prove very difficult. One problem arises from the difficulty of supplying full coded information on a P.P.I. without cluttering up the display. Others lie in the choice of the interrogating/reply frequencies. The M.C.A. is beginning by experiments with a simple transponder, which could be operated by the pilot on request and would mark his particular blip distinctively on the P.P.I. for so long as he kept the transponder in operation. This should not clutter up the display unduly although, as against this, the identification would still have to be carried in the Controller's head or pencilled on the tube. However, it is hoped to learn a good deal more about the use and requirement for transponders when a piece of equipment has been subjected to technical and operational trials.

Another need is for the accurate determination of height. Radar height finders, currently in use, rely on measurement of angles and distance. A slight error in the angular measurement becomes quantitatively more serious as the distance increases between aircraft and height finder; and, except at fairly short ranges, the accuracies of present height finders restrict their application mainly to assisting in identification and to preventing directions for unnecessary avoiding action by aircraft that are, in fact, separated by some thousands of feet. The answer to the accuracy problem is not easy, although one idea is for a transponder which, coupled to a barometric capsule, would transmit continuous identification of height as well as identity. However, this would be a very complicated apparatus if one considers that indication of 500-ft. levels over present operating heights might require 70-80 height codes, in addition to the coding required for identification purposes. This problem is still under review, but a successful solution does not appear likely in the near future.

Apart from the need for better height finding and identification facilities, a full return from the employment of radar requires better methods of displaying and applying the information derived from it. At the moment information in an Air Traffic Control Centre dealing, for example, with a stretch of airways, is derived from two sources. First, there is what is termed the "Procedural" information, which is derived from a Flight Plan and an Air Traffic Control clearance, assigning particular altitudes and times at reporting points to each aircraft. These are checked and amended from time to time as position reports are received from the aircraft and are displayed by means of written Flight Progress Strips on a Flight Progress Board. The operation of this naturally requires a reasonable amount of light. Second, there is the radar information, at present confined to the P.P.I., which has to be viewed in obscurity since there is still no certainty of the long-promised daylight viewing tube. In the absence of an effective means of "marrying" the two types of information so that both are available to a single Controller, the present emphasis in long-distance radar work has to be on dealing with a minority of situations involving closer separation and on monitoring traffic against untoward incidents, leaving the basic separation minima as required for a procedural system, rather than in speeding up the entire traffic flow by relying primarily on information "taken off the tube." This problem requires urgent attention and it should be possible to learn a good deal more when the "procedural" Air Traffic Control Centre now at Uxbridge is transferred to London Airport and occupies the same accommodation as London Radar.

If radar cover is required over a considerable area down to the lowest airway altitudes—say 3,000-ft.—and if in addition it is required to be able to detect the presence of small aircraft, such as jet fighters, it is impracticable to provide radar cover from a single site and use has to be made of additional radar stations sited where they can give the "fill in" required. Ideally, the information derived from these satellite stations should be remoted back in order to appear instantaneously and continuously on a main display at the Air Traffic Control Centre concerned. In practice, no such facility exists, and M.C.A. experiments in "telling" information back from Supplementary Radar Control Units (S.C.R.U.'s) by telephone to a main radar station is that, by the time the information has been displayed in the main Centre, it is too old. Naturally, in this context a delay of even a few minutes is unacceptable, bearing in mind the speed of contemporary aircraft. Until such time as it is possible to transmit information in such a way that it is displayed continuously, with identification and height, and with a delay of certainly no more than half a minute, the S.C.R.U.'s have to exercise separate control of particular aircraft within their own assigned airspace. This is less satisfactory than a main Control Centre, partly because of expense in staff and communications facilities, and partly because it is less satisfactory for pilots to have to pass through a multiplicity of Controls then to remain under a single Control for a longer period.

2.5. *Radar in the Terminal Area*

It has for long been M.C.A. policy to install surveillance radar at all the more important aerodromes where the density of air traffic warrants it. Originally, surveillance radar was thought of only as a component of the Ground Controlled Approach system, for use in "putting on" aircraft before handing them over to the precision element ("talk down"). As long ago as 1948, however, the United Kingdom began using the search element of the Federal GCA equipment at London and Northolt for taking aircraft from the holding patterns based on radio ranges direct to Final Approach, and the numbers of such sequences under Instrument Flight Rule conditions have already passed the 15,000 and 10,000 marks at Northolt and London respectively. The functions of surveillance radar for Approach Control may be defined as:—

(1) To give aircraft navigational assistance in positioning themselves for Final Approach.

(2) To monitor the terminal airspace against "intruders."

(3) To marshall aircraft from Holding Patterns to Final in order to secure a well-spaced traffic flow.

(4) To monitor aircraft on their take-off paths in order to reduce the separation required between aircraft outbound from an aerodrome and those in the Holding Pattern or on Final Approach.

Judged by this, the present Federal equipment has serious limitations in that its vertical cover is cut off at about 4,000 ft. and it is subject to interference from Permanent Echoes. The I.C.A.O. specification for what is termed Surveillance Radar Element (S.R.E.) was based on the old conception that this should be primarily an adjunct to the radar approach system. The cover specified is not adequate for the application of this equipment to Air Traffic Control, either in altitude or range. The specifications laid down for new British radar equipment (Approach Control Radar) went further than this and provided for cover to extend up to 10,000 ft. at about 20 miles. It also includes Permanent Echo Cancellation and there is little doubt that equipment similar to or better than this will be essential for improved landing rates in future.

At the shorter distances arising in the terminal area, the inaccuracies caused by errors in angular height finders are not so serious, but there remain the problems of identification and display. At the moment it is assumed that ultimately either some form of "daylight viewing" radar tube will be produced or that the intensities of lighting in different parts of the Control room will be so varied that the Approach, Aerodrome, and Radar Control will be integrated in the same accommodation. Since the new Control accommodation to be erected at London Airport, for example, will obviously have to be usable for many years, these assumptions have serious repercussions. Nevertheless, it is pretty clear that these Controls must be integrated for ultimate efficiency and that the display problems mentioned will have to be overcome.

2.6. *Increasing the Landing Rate*

It is generally accepted that the minimum safe separation for aircraft in a vertical holding pattern is 1,000 ft. This was confirmed by radar checks carried out last year, in which errors of some hundreds of feet were observed, often caused by pilots descending below their new altitude before levelling out. The risk of overshooting the next lower altitude is generally greater at high rates of descent, and for this and other reasons including avoiding rapid changes of pressure in unpressurized aircraft or an uncomfortable floor angle, the chosen rate of descent is about 500 ft. per minute. This means that each change of a 1,000-ft. altitude level will take two minutes, and this irreducible element in the stack-leaving interval may ultimately prove the limiting factor in speeding up movement rates from a single stack.

Investigations are therefore proceeding into the practicability of feeding aircraft to a single runway from two stacks. If successful, the same principle could, of course, be used for feeding aircraft concurrently to parallel runways. However, when aircraft have to be phased from two stacks through a common gate, preliminary theoretical studies show that, if rapid landing rates are to be attained, present standards of time-keeping must be greatly improved. Briefly, observations have shown that even for the flight from stack to gate, which may occupy about 6½ minutes, the standard deviation of flight time is as much as a minute. This makes it difficult to achieve a properly separated frequency through the gate, even from a single traffic stream, if the landing intervals are to be as low as, say, three minutes. It becomes very much more difficult when the timing of an aircraft has to be adjusted not only to preceding aircraft in the same stream but also to other aircraft that are themselves under adjustment in the neighbouring stream from the other stack. It is possible that a brilliant radar controller could achieve the desired result by his own personal knowledge and skill, subject to the fallibility of human effort; but, since no public service could be organized on the assumption that all its officers were artists of the highest order and were at the height of their powers for every minute of every watch, other means must be sought, and it is intended to carry out experiments and practical trials to this end.

In the meantime two things are clear. First, that accurate radar surveillance will be essential. Second, that some form of computor will be required to calculate the separation which will occur at the gate between aircraft on two different and curved flight paths, and if possible to display to the Controller the heading instructions that should be passed to the aircraft concerned if adjusting action is required.

It has been stated earlier that present errors in maintaining stack-leaving times, based on use of radio range and markers, may often be measured in terms of minutes; and that it is hoped to reduce this considerably by the provision of improved navigational aids. An improved navigational aid should help in accelerating the rate of departures at peak hours. Under existing systems, the nature of the facilities and the shortage of frequencies for additional installations requires departing aircraft to be routed through the same points as aircraft incoming from the airways. This means that the bottom of the stack has to be kept sufficiently high for outgoing aircraft to be routed underneath. As the stack has to be sufficiently far away from the aerodrome for the aircraft to have time to lose height before final, the longer that this intermediate flight takes, the greater the chance of variations in time-keeping. It also means that there are only about a couple of outgoing altitudes for departing aircraft and, during peak departure periods, aircraft risk being delayed on the ground for lack of a free outbound altitude at the time required. Given a flexible navigational system providing continuous indication of position, outgoing aircraft can join the airways at any predetermined point and need not suffer the present difficulties. The problem will still remain of feeding them into the airways at their required airways altitudes, but it can then be greatly reduced by radar surveillance.

2.7. *Instrument Approaches*

The combination of aids that exist or will be installed at first-class aerodromes consists of ILS glide path and localizer facilities, together with two or three marker beacons of which one may be a non-directional M.F. Beacon; GCA for use either as the primary approach aid or for monitoring an approach made on ILS; and the really first-class approach lighting system developed by M.C.A. and the R.A.E., and usually referred to as the Calvert system, after its gifted designer at R.A.E. Given some instrument similar to the Zero Reader in the cockpit (see sect. 2.2.4) the air traffic control problem, caused by baulked approaches, should be very greatly reduced. This combination of aids appears to go about as far as can be expected in providing facilities short of those suitable for enabling the aircraft actually to touch down and complete its landing run on instruments.

The step from a combination of aids that enables an aircraft to approach safely down to, say, 200 ft. after which landings are completed visually, to an aid that enables the pilot to land his aircraft and finish the landing run without taking his eyes from his instruments, is a very large one indeed. The justification for it requires careful scrutiny, from the aspects of the inconvenience and dislocation caused to passengers by being diverted, the cost and trouble to operators in having to organize handling the diverted aircraft and passengers, and the expense to administrations in having to provide diversion aerodromes and facilities. One method of bridging the gap is in the provision of thermal fog dispersal (FIDO). Assuming this is not always an economic proposition, it is worth considering what an enormous responsibility will be placed on the designers, manufacturers and operators of radio or radar facilities designed to provide completely blind landings.

The present concept of a minimum break-off height during an instrument approach, after which the aircraft should overshoot and climb away if the visibility from that height is insufficient for a visual landing, is based partly on the aerodrome terrain concerned, partly on the navigational aids, and partly on the performance of the aircraft when overshooting from such an altitude. It follows that the risks involved in overshooting on instruments from altitudes lower than this optimum break-off height naturally increase the lower the aircraft comes, since the margin for some failure on the part of the aircraft or error on the part of the pilot is progressively reduced. A radio aid designed to bring a pilot right down on to the runway and to keep him on it after touch-down, still on instruments, must therefore be able to guarantee that the aircraft with its 50 or 60 passengers will be able to complete its landings every time, and that the risk of a baulked landing, either from pilot error or from malfunctioning of the radio aid, will be negligible. The development of such aids to the point when they could be used regularly by civil passenger-carrying aircraft, and when the airborne equipment will be sufficiently easily maintained and reliable, is clearly a big and a long task. With the limited resources open to civil aviation at present, it could not be put high up on the priority list without sacrificing some other item of development. For what it is worth, my personal opinion is that, however desirable blind landings may be, at the moment the air transport industry would benefit more if priority were given to developing facilities and procedures for reducing I.F.R. landing intervals down to the order of 1½ or 2 minutes. Two-minute landing rates have been achieved from time to time already, both in the United States and the United Kingdom, but their achievement over a limited period is not quite the same thing as maintaining them regularly over a long period whenever required. An average landing interval of two minutes must mean that on many occasions the landing interval will be a good deal less, and this means working to much finer limits than might appear necessary at first sight.

A problem that confronts most Administrations at some time or another is whether instrument landing facilities, such as ILS, are justified by the traffic at particular aerodromes. Both ILS and GCA are very expensive. There are instances where the density of traffic at a particular aerodrome does not really justify these equipments, but where the air services concerned —which may only amount to six or eight movements a day—are of particular importance to the local communities. Depending on local conditions, one measure has been to install non-directional M.F. Beacons on which the aircraft could make instrument let-downs with the use of A.D.F. Unfortunately, the congestion on the M.F. band is too great to permit the number of installations desirable and other means must be sought. A promising means of meeting the requirement for an inexpensive cloud breaking and instrument approach aid is a ground installation combining a simple search radar on 3 cm, using an "A" scope, with V.H.F. direction finding. The addition of distance measuring information to the ordinary D/F let-down or QGH procedure should enable aircraft to position themselves sufficiently accurately to complete their approaches in visibility conditions considerably lower than those that will be suitable for a D/F let-down and approach only, though naturally not so low as those achievable by the use of ILS or full GCA. A model of such an aid has already been produced as a private venture and, with certain modifications, it seems promising for the purpose outlined previously.

2.8. *Special Aids to Air Traffic Control*

The continuing development of Air Traffic Control requires a number of new aids for specialized functions. The needs arise sometimes because the work cannot be performed manually at all, and sometimes because the increasing complexity of situations caused by increased air routes, more numerous movements, and faster aircraft, requires problems to be solved more quickly than the human brain is capable of doing by itself. Three or four examples will be given as an instance of the field to be covered.

It is well known that there is an inverted analogy between the problems of Air Traffic Control and those of intercepting a hostile aircraft, the inversion arising from the fact that Air Traffic Control clearances are based on calculations leading to the assumption that particular clearance concerned will lead to an *avoidance* of other aircraft by a required margin. A computor is needed to work out the take-off clearance for an aircraft in order to avoid conflict with aircraft inbound or taking off from other aerodromes at about the same time. This means that information must be fed in, derived from previous Flight Plans and in-flight information, and the computor must either provide the details of the safe clearance for the next flight or must indicate whether a proposed clearance will be dangerous or safe when details of this clearance are in turn fed into it.

The requirement quickly becomes more complicated. It will be apparent that the information fed into such a computor does not remain valid for long but must be kept up to date. Aircraft do not keep to their Flight Plans; sometimes they change them involuntarily and sometimes by request. This means that the flow of information to the computor must be continuous. It is unlikely that it will be practicable to feed in this information manually, from information received by R/T during flight, and the alternative is to provide it continuously, probably by radar surveillance. Such a computor, which would amount, in fact, to a basic component of an automatic Air Traffic Control system, would have to be introduced as part of a system and might well be difficult to introduce "on its own" into existing systems. Nevertheless, the ultimate need remains, even though meeting it may be very expensive indeed. So far as civil aviation is concerned, this need has received less publicity in the United Kingdom than it has done in the United States through the medium of the S.C. 31 Report; but obviously it is not of interest to civil aviation alone.

Another need is for a form of automatic communication between certain control centres and approach controls. If we consider the circumstances of an Airways Controller clearing aircraft to a stack under the control of an Approach Controller, it is clear that the Centre must know what is the lowest altitude in the stack to which an aircraft can be cleared from the airway. When traffic movements are not too numerous, it is possible to work to a general rule, such as that down to a certain height all altitudes in the stack will be allocated by the Centre, and below that height they will be allocated by Approach Control. However, as soon as the movement rate builds up, it becomes more than ever important that aircraft from the Airways should reach the stack ot the lowest free altitude in order to reduce their transit times to the runway. At present the information is obtained by telephone, but this adds to the noise level in control rooms, increases the strain on the A.T.C.O.s, and has in it the risk of error, in so far as any human organization is fallible. One American system—admittedly not radio but electrical—is known as the G.R.S. interlock system and consists of an apparatus with displays in each control room concerned showing altitudes in the stack, with three coloured lamps opposite each altitude. A controller wishing to obtain an altitude for which no light is showing presses a button which "interrogates" the display at the other end and, if no reservation has been made at the other end, indicates that the altitude is reserved at the transmitting end. The reserved altitude at one end is shown with a steady amber light and at the other end by a flashing amber light. An occupied altitude at one end is shown by a green light and at the other end by a red light. A small pilot model has been made and, so far as the U.K. is concerned, an investigation is being carried out by the M.C.A. Air Traffic Control Experimental Unit to observe and analyse the sequence of events in the control rooms concerned in order to get more accurate information on the requirement.

A third need is the production of simulators and recorders for training and research purposes. In training GCA staff, for example, the provision of aircraft is a very costly item and training courses are liable at times to the difficulties associated with any activities based on flying programmes. A GCA simulator is, therefore, being produced, in which blips may be placed and moved on the P.P.I. tube in accordance with data inserted to simulate the movements of aircraft. This can save a very great deal of time in that it can be used to teach the GCA pupil procedures and "patter" in the same way that pilots learn procedure and phraseology from exercises in the Link Trainer; and untoward situations can be simulated in order to observe the reactions of the pupil and to train him in dealing with such situations should they occur when he is actually controlling aircraft. When used for research into new control methods, it is essential that the simulated movements of aircraft should be consistent with those of actual aircraft.

The recorders consist both of speech recorders and of recorders of displays for GCA approaches. One problem with the R/T recorders was the injection of time signals using the morse code. Records of displays during GCA approaches could be used partly for training, partly as a check on the efficiency of the controllers or the pilots concerned, and partly, if required, for the investigation of an accident. No display recorders are yet available in this country, but the probable method will be by a continuous photographic record of the radar display, which could be integrated with the R/T record.

3. Navigational Aids over Sparsely Inhabited Territories

It is now time to consider the navigational requirements for flights over long stretches of water or sparsely inhabited territories. To get an idea of the distances involved on longer stages, the South Atlantic route between Dakar and Natal is something under 2,000 nautical miles, the Colombo-Singapore stage is nearly 1,500 nautical miles, and the stage across the Sahara from Tripoli to Kano is about 1,250 nautical miles. The drawback to astro navigation lies partly in its dependence on the weather, and partly in the fact that the positions take a little while to obtain and are not sufficiently accurate to provide a very frequent means of checking ground speeds. The provision of a more accurate radio or radar navigational aid is required not only to meet this shortcoming and to reduce the strain on the crew, but also to provide a better means of checking the progress of flight at very high altitudes. The tendency with both piston-engined aircraft and jets is to increase their operating altitudes, and it is known that, at great heights, winds may be encountered having a very great velocity but confined to a comparatively narrow stream. On world routes generally, the organization for providing accurate forecasts of upper air conditions is comparatively scanty and its augmentation is bound to be costly and comparatively slow. The quickest means of enabling an aircraft to check on its wind forecast, and to take avoiding action if it encountered one of these "jet streams," would be to provide it with a continuous indication of position and ground speed.

No existing civil long-distance aid is suitable for general application on the main routes. On some routes the noise level caused by atmospheric interference would make dependence on Consol undesirable. On others, the range of Loran, particularly overland, would be inadequate unless installations were increased to a very uneconomic scale. There are certain routes where neither Loran nor Consol would be satisfactory, either because there is nowhere to site some of the stations required for giving fixing cover, or because such sites would be in jungles or mountainous territory that would have particular difficulties of their own. We do not yet know what the solution to this problem will be, and the matter is under continuing study by the Member States of I.C.A.O.

The navigational problems on the short routes or around the aerodromes of many overseas territories are different from those previously reviewed in connection with the very busy air terminals in continents such as Europe. At certain places abroad, notably Hong Kong, the requirements are stringent because the air traffic is fairly dense and the local territory extremely difficult; but at most places the traffic movements are fairly light and the basic navigational problem remains the original one of navigating from one aerodrome to another, rather than of having to comply accurately with a detailed clearance from Air Traffic Control. Although good service is obtainable from such bearing aids as the non-directional beacon, in considering long-term improvements, it is desirable to think in terms of a position-indicating system, particularly since in some territories a forced landing could have serious consequences if the whereabouts of the aircraft were not quickly known. Apart from transit traffic by long-distance aircraft, the aircraft will in general be of the medium and smaller types, and the equipment should be as light as practicable. The ground installations should not be too complicated for ultimate maintenance by local staff under suitable supervision, and for administrative and security reasons should be capable of being sited on or close to the aerodromes served or to easily supervised territory.

Looked at from this point of view, there are obvious attractions in point source aids like VOR/DME, sited on the aerodromes concerned. However, if even the major operators in the wealthiest parts of the world are having second thoughts about the economics of completely equipping their aircraft with this system, it is even less likely that overseas operators of smaller aircraft will use such equipment. The conventional Decca Chain, as known in Europe, suffers from the administrative disadvantage in such areas of requiring its master and slave stations to be sited some distance apart which, in certain territories, could cause difficulties. However, an application of Decca, known as the Integrated Track Range, was tried out with success in this country earlier this year, and its stations, which consisted of two about three miles apart and one at a distance of about 200 miles, could be sited on or as close to aerodromes as flight path clearances would permit. Aircraft equipped with ordinary Decca receivers and the Flight Log, or even without the Flight Log, could obtain position information over a large area, which would come in extremely useful when circumnavigating storms or making special flights, as well as during normal operations. The equipment would be no less useful to transit aircraft, coming from more populous areas where Decca was the basic aid.

At many places abroad, ILS is required either because of the frequency with which low visibilities occur, or because if aircraft have to carry enough fuel to go to an alternate aerodrome, the reduction of pay load becomes excessive because of the distance of the alternate.

At other places, such as up-country aerodromes, there is an obvious attraction in the idea of the simple radar/DF system mentioned earlier. (Sect. 2.6.)

The need for a completely airborne type of navigational aid arises not so much from conditions on the ground as from the weather conditions encountered over some of these territories. In certain seasons, and over large parts of some routes, dense cloud extends to a great height, storms are frequent, and the turbulence encountered can be very uncomfortable to passengers, and could be worse than uncomfortable if flown into at a very high speed. An Airborne Search Radar, formerly known as "Cloud and Collision Warning," has been developed to the requirements of M.C.A. and B.O.A.C. and, following extensive trials sponsored by M.C.A. and carried out by B.O.A.C. in a Hythe flying boat and a Viking landplane, both in this country and in South-East Asia, it has been adopted by B.O.A.C. for use on some routes. The equipment consists of a primary airborne search radar giving a P.P.I. display and operating on a wavelength of three centimetres. It provides an indication of the direction and distance of the areas of main turbulence, which are generally, though not always, airspaces in which the moisture suspended in the up-currents of air is in larger droplets, giving a brighter return on the radar tube. Flight records, and photographs of the displays on the tube when approaching cloud formations, provide a striking illustration of the way aircraft may be manœuvred round or between areas of turbulence even though the pilot may be flying on instruments all the time. An additional advantage of this aid is that it provides a degree of "ground painting," which should be of considerable value over mountainous regions and on those routes with occasional landmarks such as coastlines, islands, or large rivers, whose presence cannot be detected visually through cloud.

A general need, which is not confined to operations in sparsely inhabited territories, is for a cheap form of Terrain Clearance Indicator. Again and again crashed aircraft have been found within a few feet of the summit of the mountain they hit in flight, and in many instances the pilot might have been warned in time to take avoiding action if his aircraft had carried some form of Terrain Clearance Indicator. It is often stated that, generally speaking, a crash into the top of a hill is the last of a series of errors and that if proper use had been made of facilities available earlier, or if good judgment had been exercised, the aircraft need not have got into a position at which a crash was risked. Also, the types of radio or radar altimeter suggested previously for Terrain Clearance Indicators are expensive and sometimes raise installation problems as well, so that their general adoption is slow If, however, some much cheaper form of T.C.I. can be produced, which would not need to give an accurate and continuance indication of height above terrain so much as an indication of when the aircraft was within a preset minimum clearance, it might well be adopted on a wide scale. Some work has been done in this direction, but no equipment to meet this requirement has yet been developed to the stage when it can be demonstrated in flight.

4. Communications

The need for faster fixed communications obviously increases with the speed of aircraft, since a departure signal to Company or Air Traffic Control at destination is not of much value unless it reaches them well before the aircraft comes within their sphere of action. This, however, is more a matter for the deployment and application of existing types of facilities and for improvements in procedures than for the development of entirely new equipment. Dealing with mobile communications, the congestion already experienced on V.H.F. is being reduced by a steady programme for increasing the number of frequencies in use, coupled with a constant drive to improve the standard of R/T phraseology. However, the possibility has been raised that, when Air Traffic movements exceed a particular density, the method of communicating by R/T will prove too slow and cumbersome in the Terminal Area and will ultimately limit the movement rate. The methods proposed for replacing R/T are some form of symbolic display operated in the cockpit by signals transmitted from the ground, and vice versa. It is not yet clear whether a sufficiently wide range of instructions could be provided in a display that was not too cumbersome, nor how one could check that the correct information had been received or obtain a timed record of what has been passed and received. Another method, suggested in the American Report S.C.31, is what is known as the "Private Line" system. This was visualized as a component in an Automatic Air Traffic Control System which would communicate with aircraft in selected blocks of airspace without knowing their identity and without requiring the co-operation of their navigational equipment, would transmit information derived from the aircraft to other elements in the sequence of automatic control equipments, and would transmit back to the aircraft clearances obtained automatically.

Estimates of probable traffic densities in this country, R/T frequencies available, and the minimum to which essential R/T communications could be cut in the landing sequence, lead to the conclusion that the use of R/T should not prove the limiting factor in the movement rate for a very long time, and that priority needs to be given to the development of other equipments first. This statement should not be taken to imply complacency, but as a rough assessment of the orders of priority, taking into account the limited development effort available. For en route communications, even on long stages, there is an increasing tendency to use R/T on High Frequency, rather than W/T as formerly. This is mainly because of the flexibility it offers in the exercise of Air Traffic Control and of Operational Control by the Company concerned, and also for its application to an improved in-flight Met. Service. It is not proposed to elaborate further on this here because, once again, it is more a question of the application of known equipment and techniques rather than of developing something entirely new.

5. Requirements of Turbine-engined Aircraft

Too often the opinion is expressed that the turbine-engined aeroplane is such a difficult kind of aeroplane that it cannot possibly be operated with present types of ground organization. It is true it is not yet known what kind of ground organization will be needed when the majority of aircraft will be driven by gas turbines, but the needs of the next few years can already be foreseen in part, and information available so far does not yet substantiate this opinion. For over a year the Ministry of Civil Aviation has been examining the requirements for the successful introduction of turbine-engined aircraft in conjunction with the two British Corporations, the manufacturers concerned with the types of turbine aircraft to be operated, and such representative organizations as the British Airline Pilots Association. This work has culminated in a series of flight trials with a Comet lent to B.O.A.C. this summer and it is fairly safe to say that, given a period in which the numbers of jet aircraft will increase steadily rather than by leaps and bounds, they raise few problems peculiar to themselves but serve rather to highlight difficulties under which piston-engined aircraft are already labouring in existing systems.

The main characteristics of turbine-engined aircraft relevant here are their prolonged periods of climb and descent, the very much higher cruising speeds of turbo-jet aeroplanes as compared with piston-engined types, their comparatively heavy consumption of fuel at low altitudes, and, arising partly from this, the desirability of their being able to commence their descent from cruising altitude at an accurately determined distance from destination to give them an economic descent path.

The problems raised by their prolonged ascents after take-off and descents before holding altitude have been outlined earlier, and it has been pointed out that the navigational system of radar surveillance needed to facilitate this is also needed for piston-engined aircraft of higher performances operating to-day.

One decision called for in the operation of all high-altitude transport aircraft is when to begin the descent from cruising altitude. If it is begun too late, an extended let-down will be required near or over the destination, with consequent loss of time and efficiency. If it is begun too early, the fuel consumption of the jet will be increased by the unnecessary time spent at low altitudes. Present thinking favours a gradual descent from cruising altitude beginning about 150 miles from destination and proceeding at normal rates down to about 20,000 ft., at which altitude the cabin pressure will be equivalent to sea level. A more rapid descent may be made through the last 20,000 ft., the rates of descent varying over fairly wide limits, but dependent to some extent on a comfortable floor angle for the passengers. At the moment, appropriate navigational aids for the initial positioning would be Decca or, in sparsely inhabited areas, the Decca Integrated Track Range referred to previously, or Distance Measuring Equipment on 200 Mc/s or 1,000 Mc/s. It may be mentioned that the M.C.A. has already sponsored the temporary installation of prototype DME at Rome and at Cairo in order to obtain experience of its efficacy when used by B.O.A.C.'s Comet aircraft. The heading facility, obtainable on

British DME receivers, is regarded as a useful additional asset.

The need for rapid fixed communications to serve flights by very fast aircraft has been mentioned previously. On the air/ground side, the range of communications of Very High Frequencies will be extended considerably at jet cruising altitudes with a corresponding increase in the area of interference. As an interim measure, it is intended to use special frequencies in the United Kingdom and Western Europe, and it is expected that the problem will be reduced when additional frequencies are brought into operation in the future.

For efficiency and economy, diversions, because of weather, should be decided on before the aircraft has descended from its cruising altitude. This means that a Flight Watch on meteorological conditions at the destination aerodrome and its alternates should be maintained by the main Meteorological Office concerned throughout the flight, and it should be practicable to get the latest Met. information through to the aircraft as quickly and as often as required. In addition, a Flight Watch will probably be maintained by the operators in order to pass operational information and advice to the aircraft in flight. Both these requirements point to the desirability of long-distance radiotelephony facilities on High Frequencies—a need which has already been recognized and met on certain routes operated by piston-engined aircraft. Finally, the need for information on conditions in the upper atmosphere, particularly winds and temperatures, is important; and this will be reflected in the need for more "Rawinsonde" stations.

6. The American Approach

In the space available only limited reference is possible to American thinking on these general problems, but this paper would be incomplete if it did not refer to the celebrated S.C. 31 Report. Arising from a request by the President's Air Co-ordinating Committee to the Radio Technical Commission for Aeronautics in 1947, the latter established a Special Committee No. 31 "For the purpose of developing telecommunications for the safe control of expanding air traffic." The report appeared in 1948 and its preparation required the efforts of 85 experts and cost about $1,000,000. The grandiose scale of its recommendations was staggering. The Committee took as one basic need the drafting of a "common plan" whereby all civil and military users in certain defined airspaces would conform to the same Flight Procedures and would use the same kind of navigational and other equipment for doing so. Put very simply, its objective was to raise movement rates under all weather conditions to those obtainable under Visual Flight Rule conditions, which can be very high indeed when separation in the landing sequence is largly left to the continuous visual information available to the pilot. Among the principles required to achieve this objective, S.C. 31 laid down that information to be used in traffic control should be derived from ground equipment, whereas information to be used in the aircraft should be derived from airborne equipment; that safe separation of aircraft should be provided by automatic means and the flow of traffic should be controlled automatically; and that human operations should be reduced by mechanical or electronic means whenever possible. The latter devices were to be interlocked, so that they could not be misused or automatically set up a hazardous situation, and to operate on the "closed circuit" principle with "fail safe" indications of malfunctioning. The system should indicate when and where aircraft could be inserted into an existing traffic flow and should automatically transmit to the aircraft and display before the pilot traffic control clearances. These are but a selection from the requirements specified and they were to be met by some 20 different classes of equipment, including monitoring equipments. The Committee visualized that achievement of the Ultimate Plan would be via a Transition Plan. The cost was estimated at over $1,100,000,000.

These stupendous proposals were arrived at by the logical process of defining ideal ultimate requirements regardless of other considerations, and then setting down the nature of the equipments required to produce these results in conditions in which operation by human elements was considered to be too slow and too fallible. There are signs, however, that realization of the tremendous cost and effort involved in relation to the finances of the air transport industry, and the changed situation in the United States as the result of rearmament have led to the time scale in S.C. 31 being changed, and apparently some of the basic assumptions also. From such subsequent reports as that of the Operational Policy Group of the Air Co-ordinating Committee, issued last December, it looks as though the United States is now prepared to spend a much longer time in a Transition Phase and to take the experience gained during this period more into account when balancing the requirements and practicability of the Ultimate Plan. In particular, it looks as though they now hope to make use of VOR for a much longer time and are resigned to the probability that for economic reasons they will not be able to rely on most aircraft being equipped with DME for several years. They are as convinced as we are of the value of Surveillance Radar in the terminal areas and are considering its application along their airways. Like ourselves they are pressing ahead with the development of transponders suitable for civil aircraft for identification purposes, and are providing ILS and GCA for instrument approaches. The most noteworthy differences of opinion between American and British thinking on these problems is, firstly, that the U.K. believes VOR will prove inadequate for the densities and types of traffic to be encountered in a few years' time; secondly, that the U.K. foresees the growth of cross-country helicopter services, and the consequent need to provide navigational aids for them, whereas some Americans at least consider that helicopters will only be used for local services serving districts around and in large cities.

7. The Principle of Universal Standardization of Aids

This principle was considered axiomatic a few years ago, and was undoubtedly very attractive in the post-war period when trunk route operators were obliged to equip their aircraft with a far too varied selection of navigational and communications facilities. Its rigid adoption would, however, raise considerable technical and economic difficulties. Meetings to discuss standardization of aids, such as those of the Divisions of I.C.A.O., naturally tend to think of areas where the problems are most acute and to specify aids with those particular areas in mind. Experience has shown, however, that insistence on installing thoughout a long route all the aids required at the busiest terminal on that route will make extravagant and unjustified demands for the provision of facilities at some staging points where the traffic is very much less and the general weather conditions may be very much better. Furthermore, at such places the needs of the local operators of feeder lines may be for equipment that can be carried in small aircraft and is lighter and less bulky than the equipment that can be accommodated in the large long-range aircraft. In these circumstances, the aids supplied at the small aerodromes should be usable by aircraft carrying the more comprehensive equipments without necessarily emulating the full scale of services provided in really busy areas.

The exact selection must depend on the local conditions, and also on the proportion of traffic coming from long-haul operators, but the objective should be to decide on what is to be the highest common factor of equipment along the routes. It is, of course, arguable that it is a hardship for the long-haul operator to carry a comprehensive range of navigational equipment for a long way and at some cost in space and payload if, for a large part of a route, he cannot utilize much of his equipment for improving the efficiency of his operations. On some routes, however, the only bad weather areas where traffic is really dense exist at the ends of the route. In these conditions, it may be practicable for an aircraft to off-load some of its equipment at the first stopping place after a busy area, and to pick it up ready for use at the same point on its way back and before entering the congested region. Such a procedure is already in use by some ocean-going ships for the highly accurate navigational aid employed when close to certain coastlines and for entering harbours, but not when crossing the oceans.

8. Conclusion

If this paper has dwelt on difficulties and imperfections rather than on the considerable results being obtained in air traffic control and navigation with existing equipment, it is only because it appeared that the main interest at this session would be focused on future developments rather than present achievements. There is no doubt that equipments and procedures in use to-day enable regular services to be maintained as a routine in weather conditions that, before the war, would have confined operations to the occasional exploits of "intrepid birdmen." The emphasis on shortcomings has only been made here because one of our slyest temptations is complacency and, for those close to the problem, the rate of progress should always seem maddeningly slow in relation to the ideal objective.

DISCUSSION

C. J. Carter: With reference to the use of search radar for Terrain "Clearance Indication," as Mr. Stallibrass says: "It is often stated that . . . a crash into the top of a hill is the last of a series of errors . . . and the aircraft need not have got into a position at which a crash was risked." I have heard it said that a radar terrain clearance indicator is unnecessary because the pilot knows his position and so knows his altitude. Yet aircraft *do* fly into hillsides.

'It seems to be overlooked that an aircraft does not fly into a hillside because of a faulty altimeter, but because the pilot is lost—in a greater or less degree. It is when a pilot is lost that terrain clearance becomes so valuable an aid.

An American operating company made a practice over years of flying an aircraft, equipped as the Radar Terrain Clearance Indication to the location of every accident in the U.S. where an aircraft had flown into the side of a hill. I believe that in every case it was able to demonstrate that a Radar Terrain Clearance Indicator would have prevented the accident.

These indicators are, as Mr. Stallibrass says, expensive and may be difficult to install, but, in my view, the value of preventing one particular fairly common type of fatal accident is so great that their cost and their installation problems are a price which should be paid.

J. H. Evans (*Associate Member*): At least one manufacturer is very alive to the need for a transponder and is hopeful of demonstrating a solution in the very near future, possibly as early as the end of the year.

I also think that the problem of traffic control will be much eased by the introduction of A.C.R. Mk. VI (Surveillance Radar). This has already been demonstrated and the prototype version is expected to be working in August, 1951.

The solution to the problem of remote display has already been demonstrated. At least three methods are available and no doubt can be applied when required.

J. N. Anslow: Mr. Stallibrass has outlined the navigational requirements and position-fixing accuracy necessary by modern aircraft to achieve a landing rate greater than one every three minutes at a main terminal airport. He has shown that such an accuracy is required to enable the aircraft to inform the ground control of its position, and hence the aircraft may be given certain manœuvres in order that its arrival at an airfield may be accurately predicted. Assuming that, from an air operator's point of view, existing position-fixing devices are adequate for normal flight operation and navigation, the carriage of additional aids to enable the accuracy outlined to be obtained will be an economic burden to the airline operator. This is specially true if the apparatus is useless for much of the en-route navigation and is only required at two or three of the world's major air terminals.

Therefore, I would suggest that accurate control and position-finding of aircraft should be undertaken entirely by the ground control organization. A radar reporting system might be considered, whereby aircraft on crossing the coast come under strict radar control in a manner similar to that in which the final stages of aircraft approach are controlled by G.C.A. In this way, the pilot can monitor the ground controller's instructions by his conventional navigational aids, such as A.D.F.s, fan markers, etc., and use his existing V.H.F. communications for this "talk-you-along" facility. No high-accuracy apparatus need be carried additionally by the aircraft, except a radar responder with perhaps an "on-request" identity coding system. In this way, the unnecessary phase of ground control, having to ask the pilot where he is before movement instructions can be given, is removed.

AUTHOR'S REPLY

I agree with Mr. Carter that on most, although not on all, occasions when aircraft crash into hillsides, it is because they are lost rather than because their altimeter is faulty. If one considers the chain of events that led up to a crash, however, it may well be that the overall reason will be reduced more by concentration on crew training and the provision of suitable navigational aids than by the installation of a very expensive radar altimeter. It is a question of deciding where the link is weakest and how much it is practicable to go on strengthening it. Faced with these considerations, I think that the first requirement is the provision of good navigational facilities. At the same time there is no doubt that there would be very strong justification for a terrain clearance indicator if one could be produced without prohibitive expense.

It is true that good work is being done on aircraft transponders and, like Mr. Evans, we hope that their simple application will be demonstrated without too great a delay. Radar to the A.C.R. Mk. VI specification will also make a big difference and all I will say here is that we have been waiting eagerly for A.C.R. Mk. VI for a long time and shall be very glad when we get it.

As far as remote radar displays are concerned, Mr. Evans is correct in stating that at least three methods are available, but none of those that we have tried out so far meets the full requirements. It may be that the ultimate solution will be in an application of television.

Whether existing position-fixing devices are adequate depends on how one assesses the needs for normal flight operation and navigation. Personally I consider them inadequate, if only because aircraft continue to get lost and hit hills, even in Europe, with loss of life. It therefore seems to me that an accurate navigational aid on which an aircraft commander should be able to rely is essential.

I agree with Mr. Anslow on the important role of radar. The difficulty in the United Kingdom is that the traffic coming into London only enters our control 60-80 miles away on the busiest routes and we cannot use radar for closely spacing outbound aircraft from the U.K. if they suddenly have to change to wider spacing under a non-radar system, half-way across the Channel. Nevertheless, quite a lot can still be done within that 60-80 miles with incoming aircraft, although improved radar equipment is needed to be fully effective. One advantage of using primary radar is that implementation can be carried out directly by the U.K., without having to wait on the outcome of protracted discussions through ICAO.

NOTE ADDED BY AUTHOR—DECEMBER 1951

Since the reading of this paper at the Convention, certain parts of it have become out of date. In particular, the author considers that Section 2.8 on "Special Aids to Air Traffic Control" needs some qualification.

A very much simpler computor is required to start with, and the need and specifications for the more complicated one cannot be determined until sufficient experience has been gained with the simpler ones.

SCANNING AND E.H.T. CIRCUITS FOR WIDE-ANGLE PICTURE TUBES*

by

Emlyn Jones, B.Sc.†

A Paper presented at the Fifth Session of the 1951 *Radio Convention on August* 22*nd in the Cavendish Laboratory, Cambridge*

SUMMARY

One of the major factors which have contributed to the low cost of television receivers, despite the increased cost of production, is the introduction of a.c./d.c. technique. In this country this implies the rather low H.T. rail potential of about 190 V. The demand for bigger and brighter pictures, on the other hand, has led to an increase in the scanning angle of some C.R. tubes, and of the final anode potential. Both these factors imply greater scanning power. The paper is concerned with combined scanning and E.H.T. circuits developed to meet this situation.

The operation of energy-recovery scanning circuits is explained, design procedures established and typical circuits are given in detail. Particular attention is paid to the problem of obtaining satisfactory linearity, and a method is described of correcting the horizontal nonlinearity by means of a saturated reactor. Details are also given of an efficient scanning coil wound on a slotted ferrite yoke.

CONTENTS

1. Introduction

Line Timebase Generator circuits for wide-angle scanning are becoming increasingly complex. This paper attempts a simple explanation of their operation and design. Arguments for and against the use of larger angles of deflection are not discussed; one line of thought only is pursued right through to the design of a suitable circuit.

Larger television pictures imply the use of either a projection system or a large C.R. tube. Considering only directly-viewed tubes, a larger picture requires an increase in beam power if the brightness is to be maintained. Since the E.H.T. supply is commonly obtained by rectifying the flyback pulse, this involves a greater load on the scanning circuits. Consequently, these circuits should control a greater amount of power if the E.H.T. load is not to have an excessive reaction on the scanning system. It therefore seems logical to give the scanning generator more work to do, i.e. increase the angle of deflection, since this brings the advantage of a shorter C.R. tube length, which in turn enables a smaller and therefore cheaper cabinet to be used.

Whatever the reasons, the fact remains that C.R. tubes having deflection angles of 70 deg., or even 90 deg., are in extensive use in the U.S.A.

* Manuscript received July 3rd, 1951.
† Mullard Research Laboratory.
U.D.C. No. 621.397.621.

and are gradually being introduced in this country.

An increase in the scanning angle results in three separate effects: the energy required to deflect the beam is increased, the relationship between the movement of the spot and the scanning-coil current departs further from linearity, and certain electron-optical aberrations in the system increase in magnitude.

To make matters worse, the British designer has been forced, for economic reasons, to design the popular receiver with transformerless power supplies, which, if he aims at a voltage range extending down to 200 V a.c. or d.c. restricts the high-tension line to some 180 or 190 V. He is thus much worse off than his American counterpart, who usually reckons on at least a 250-V h.t. line, which may be obtained by voltage-doubling from a 110-V a.c. supply.

Our typical problem is, then, to design a scanning system, operating from a 190-V h.t. line, generating 13 to 14 kV with good regulation (5 to 7MΩ) and scanning a 70 deg. C.R. tube with some margin to spare for component tolerance and valve deterioration during life. Non-linearity is to be less than ·05.*

There are two main aspects of this problem:—

(*a*) The design of the circuit to give good linearity, adequate E.H.T. supply with good regulation, and a low power input.

(*b*) The design of the scanning coil to give the best compromise between the requirements of high sensitivity, freedom from corner-cutting, uniform focus, picture rectilinearity, and L/R ratio which affects linearity.

We shall confine our attention mainly to (*a*).

Although much has been written about (*b*), the design of scanning-coils is still largely empirical and we will content ourselves with giving details of a satisfactory coil which will provide us with the numerical data necessary as a starting-off-point in the design of the circuit.

2. Energy Considerations

Conventional C.R. tubes have scanning angles of about 63 deg., and the picture format is most usually rectangular with rounded corners, giving a maximum horizontal deflection of 59 deg.

The new tubes have a scanning angle of 70 deg. and "double-D" deflection is often employed, giving a maximum horizontal deflection of 70 deg. If the coils were of the same length this would involve an increase in the stored energy in the ratio of $\left(\frac{\sin 35°}{\sin 29{\cdot}5°}\right)^2$ or 1·37, since the energy in the field is proportional to B^2 and the required flux density (B) is proportional to the sine of the deflection angle.

However, the increased scanning angle involves a reduction in coil length if the beam is not to be "corner-cut" by the C.R.T. neck, and this reduces the sensitivity of the scanning coil. Furthermore the neck diameter of the wide angle tubes is slightly larger than that of the narrow angle tubes. The effect of these factors is to make necessary about 2·2 times the field energy in this particular comparison. Since many other factors than deflectional sensitivity enter into the design of a scanning coil, this figure of 2·2 should be regarded only as an indication of the order of magnitude involved.

3. Power Considerations

The mean power required to operate the time-base depends on what is done with the stored energy. If it is completely dissipated each cycle in resistive losses, the least amount of power P, which can operate the system will be $P = Wf$, where W is the stored energy and f the horizontal scanning frequency. As an example, we may take a typical scanning coil and transformer requiring a stored energy of 1 millijoule to scan a 70-deg. tube such as the Mullard MW 41-1 with a final anode potential of 14 kV, at a frequency of 10,125 c/s. The power required, if all this energy is dissipated each cycle, is 40·5 W; at 180 V this corresponds to a minimum mean current of 225 mA. In practice the current will be much greater than this, probably over 300 mA, because the energy conversion of the valves and circuits will be less than 100 per cent. Considering that the rest of the set will only take about 120 mA, this is by comparison rather excessive.

It is therefore highly desirable to find means of recovering some of the stored energy in a useful form.†

* Non-linearity is expressed as follows. If the picture is modulated with narrow vertical bars corresponding to equal intervals of time, and if x and y are the widths of the widest and the narrowest of these, then the non-linearity is $2(x - y)/(x + y)$.

† Distinction can be made between the "damping diode" circuit in which a portion of the scan is obtained

4. Circuits for Energy Recovery

The basic circuits for energy recovery were invented in England by Blumlein[2] as long ago as 1932. However, the extra components required were not always justified by the moderate power saving possible, when narrow angle tubes, operating at E.H.T. potentials of 5-7 kV, and silicon-iron core and yoke materials were customary.

Blumlein's basic circuits have been arranged in many forms and it is of importance to choose the most convenient one. The fundamental mode of operation is the same for them all, and for brevity we shall describe in detail only one circuit, relegating the others to Appendix I, where their advantages and disadvantages are listed.

4.1. *Basic Mode of Operation*

Figure 1 shows the circuit we propose to discuss, leaving the questions of heater, screen grid and E.H.T. supply to be considered separately. We shall also ignore the resistance of the windings and the leakage reactances of the transformer at first. The stray capacitance, however, must be included as an essential component of the system. This is shown as C1, and since the transformer is an ideal one we can connect it across any winding, say 1-4. We must assume an initial charge of E_b on capacitor C2. E_s is the potential of the H.T. rail. L_y is the scanning coil; C3 is the normal H.T. smoothing capacitor.

Suppose the grid potential of V1 is suddenly reduced from a very negative value to zero. The potential E_4 will fall to the knee voltage of V1. The potential of terminal 2 will fall in proportion to the turns ratio and the resulting potential difference between 1 and 2 will appear across L_y causing a linear positive increase of current i_y. Let us suppose that terminal 3 has been tapped into the transformer at such a point that with V1 "bottomed" the cathode potential of V2 is very slightly more positive than its anode, so that no current flows in V2. The current i_y will therefore be supplied via V1, transformed by the ratio $\frac{n_{12}}{n_{14}}$. In addition V1 will supply the magnetizing current of the transformer, the sum being i_4. All these currents will increase linearly until at a suitable instant we will suppose that the grid voltage of V1 is made very negative, so that i_4 becomes zero. The current in the transformer now flows into the stray capacitance C1 and its value i_c is initially equal to the final value of i_4. The inductive system executes slightly more than a half-cycle of oscillation with the capacitance C1, the potential E_4 becoming highly positive in the process, and then falling until held steady by the diode V2 at approximately the potential from which it started. The current i_c is proportional to the rate of change of E_4 and therefore reverses in sign during this period. When E_4 is held steady by V2, i_c suddenly falls to zero, the inductive current now being supplied by i_3. (We note that with the sign chosen i_3 must always be negative.) The magnitude of i_3 is such that it can maintain the magnetic energy of the system unchanged during the instantaneous redistribution of current. The magnitude of i_c at this instant would be equal to that at the beginning of the oscillation but of opposite sign, if the circuit had no loss. If the circuit has a loss corresponding to a finite Q, say Q_r at the frequency of the oscillation, the values at the beginning and end of the oscillation are related by the equation

$$\hat{\imath}_{c_2} = e^{\frac{-1 \cdot 65}{Q_r}} \hat{\imath}_{c_1} = \delta \, \hat{\imath}_{c_1} \quad \ldots\ldots\ldots\ldots (1)$$

$$\text{where } \delta = e^{\frac{-1 \cdot 65}{Q_r}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

using the notation of Fig. 2(a).

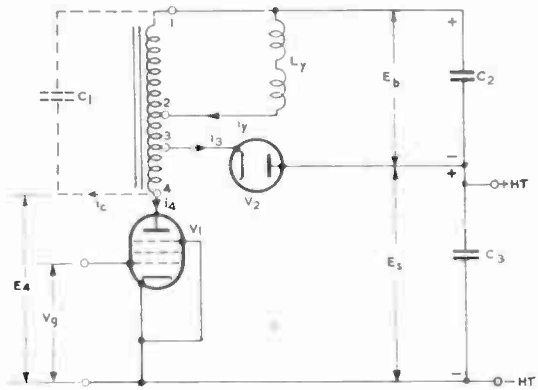


Fig. 1.—Series feedback energy-recovery circuit using auto transformer.

from a diode, but the energy is dissipated in a resistor, and the "efficiency diode" in which the diode provides a portion of the scan and in addition returns energy to the scanning valve or H.T. line. This distinction is of little importance because the damping diode is hardly ever used in modern circuits.

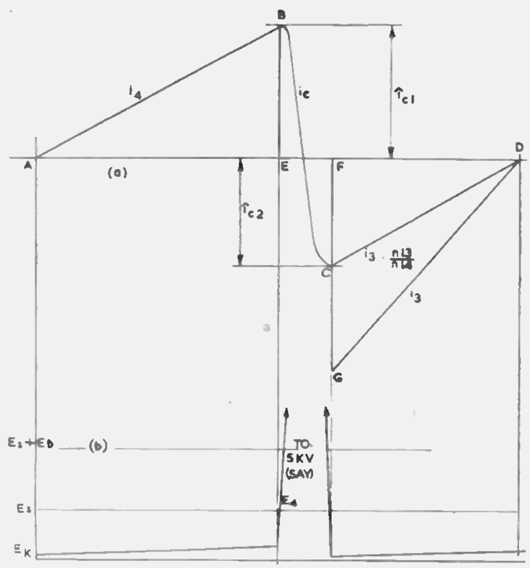


Fig. 2.—Current and voltage waveforms for the circuit of Fig. 1.

Thus we have

$$\hat{\imath}_3 = -\frac{n_{14}}{n_{13}}\,\hat{\imath}_{c2}$$

$$\text{or } \hat{\imath}_3 = -\frac{n_{14}}{n_{13}}\,\delta.\hat{\imath}_4 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

where $\hat{\imath}_3$ and $\hat{\imath}_4$ are the peak values of i_3 and i_4.

We postulated that during the first part of the cycle the potential of 3 is only slightly different from E_s. Hence the rate of change of i_y will be almost the same as before, which implies a constant velocity of the scanning spot, i.e. good linearity of scan.

At the instant D (Fig. 2*a*) the pentode grid is again made positive, and the cycle restarts as from A.

We notice that during the period A-E the current i_4 flows out of C2. During F-D the current $-i_3$ flows into C2. Provided the areas A B E and F G D are equal the potential across C2 is maintained at the value E_b which is assumed.

The peak-peak current swing required from the pentode is less than that required in a conventional system not employing a diode, in the ratio

$$i_4 \Big/ \left(i_4 + \frac{n_{13}}{n_{14}}\, i_3\right) \text{ or } 1 \Big/ (1 + \delta)$$

The mean current taken from the supply is equal to the area F G D averaged over a whole cycle A-D. This quantity of electricity F G D is raised to the potential $(E_b + E_s)$ in the process of recovering the stored energy from the magnetic circuit, and is then available at this potential for utilization by V1. In the absence of the efficiency diode the same quantity would have had to be supplied from a supply line of $(E_b + E_s)$ and would even then have performed only slightly more than half the scan (that is A B). The saving in power can thus be very considerable, the exact amount depending on the losses in the circuit. A lossless circuit would require no power at all.

The potential increment E_b by which the incoming electricity is raised depends on a number of factors with which we shall deal in detail later, but it is of interest here to note a fundamental relation between this potential and the "overall Q" of the system, defined as the ratio

$$Q = \frac{\text{Energy recovered from magnetic system and stored in C2 per cycle}}{\text{Energy dissipated per cycle.}}$$

the cycle being that corresponding to A-D. Taking the area F G D $= q$, the energy input per cycle is $q\,.\,E_s$, and this must equal the losses. The energy recovery per cycle is qE_b. The ratio Q is thus $\dfrac{q\,E_b}{q\,E_s}$

$$\text{or } E_b = Q\,E_s \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

In making the preceding simple analysis we ignored the variation of voltage-drop across V1 and V2 during the cycle. As this is of great importance in connection with linearity it is necessary to interpret Fig. 2*a* in terms of actual valve characteristics. We really have two sorts of diagram superimposed. A B E and F G D are diagrams of actual currents i_4 and i_3, and are of interest when considering the charge flowing in and out of C2. A B C D can be regarded as an ampere-turn diagram in terms of an equivalent current flowing in the winding 1-4, and is of interest in determining the magnetic effect of currents in the transformer. The first diagram can be discontinuous, the second cannot.

If we imagine a variable source of potential applied across terminals 1-4 we can plot the currents i_4 and i_3 against E_4. The current i_3 will

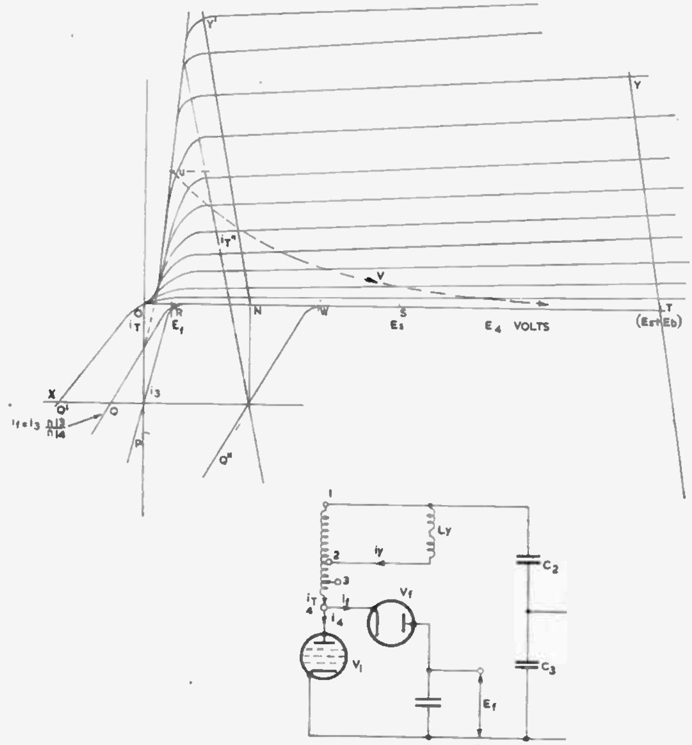


Fig. 3.—I-V diagram useful in design, and equivalent circuit postulating a fictitious diode Vf.

depend only on E_4, but i_4 will also depend on the potentials of the electrodes of V1. This gives us the diagram of Fig. 3 in which the positive region refers to the current i_4 and is the ordinary I_A-V_A diagram of the pentode V1. The negative portion refers to the current i_3 and here we have two characteristics. Curve P is the actual value of i_3 plotted against E_4, which is the same as an ordinary I_A-V_A curve for V2 but with the voltage increased in the ratio n_{14}/n_{13}. Curve Q is the same thing with the current reduced in the ratio n_{13}/n_{14}, i.e. it is the curve of a diode which, if connected between terminal 4 and a suitable source of potential E_f, would produce the same number of ampere-turns in winding 1-4 as V2 actually does produce in winding 1-3. This may seem an unnecessary complication, but the curve Q is of great help in understanding various other modes of operation than the basic one described, and the relation between these other modes and the linearity of scan. We draw the fictitious diode V_f taking a current i_f as shown in Fig. 3.

We note that the diode commences to take current when terminal 3 (Fig. 1) is at the potential E_s. Hence R (Fig. 3) is at a point such that

$$\frac{RT}{ST} = \frac{n_{14}}{n_{13}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

By varying this turns ratio, therefore, we can place curve Q anywhere we please along the horizontal axis.

The current flowing in the transformer is the sum of i_4 and i_f

or $i_T = i_4 + i_f \ldots\ldots\ldots\ldots\ldots\ldots (6)$

Since i_f is always negative, this is the difference between the curve Q and the pentode current locus.

The rate of change of i_T will satisfy the relation

$$L\,\frac{di_T}{dt} = E_s + E_b - E_4 \ldots\ldots (7)$$

where L is the inductance between terminals 1-4.

We now select a path on the $I_A V_A$ diagram (a simple and convenient one is that for $V_{g1} = 0$) and plot the difference between this and curve RQ to find a locus of i_T. A time-scale can now be plotted along the i_T locus to satisfy equation 7 at each point, and from this the resultant transformer current i_T and its components i_4 and i_f as functions of time can be obtained. In actual practice one may take $\frac{di_T}{dt}$ constant, and the time scale can be drawn linearly along the vertical axis between the peak negative and peak positive values of i_T. This is because non-linearity is corrected for as described later. The current i_y will be of the same form as i_T but balanced about the horizontal axis because there is no direct voltage around the loop 1, 2, L_y.

The presence of d.c. resistance in the scanning-coil and transformer windings will modify the picture somewhat. We can regard the change in potential across L_y due to the resistive drops to be produced by a modulation of E_b to produce the same effect with a resistanceless coil and transformer. Thus we draw through T, Fig. 3, a resistive load line at the correct slope to make the necessary adjustment. The value of this resistance can be found from an analysis of Fig. 4, which shows the equivalent circuit for either diode or pentode conduction.

We write

$$V_3 = V_L + r_y i_y$$
$$V_1 = V_3 - r_a (i_y - i_1)$$
$$V_2 = (n - 1) V_1$$
$$V = nV_1 + r_b i_1 - r_a (i_y - i_1)$$

from these we find.

$$V - nV_L = i_1[n^2 r_y + r_b + (n - 1)^2 \; r_a]$$

provided $i_y = ni_1$, i.e. we ignore magnetizing current compared with load current. This is justified in practice because it represents a small error in a small correction. If there is no resistance, $V = nV_L$. The term $n^2 r_y + r_b + (n - 1)^2 r_a$ is thus the equivalent resistance in series with the primary which is what we require.

For terminal 4 of Fig. 1 this becomes:—

$$R_4 = \left(\frac{n_{14}}{n_{12}}\right)^2 r_y + \left(\frac{n_{14}}{n_{12}} - 1\right)^2 r_{12} + r_{24}$$

For terminal 3 it is

$$R_3 = \left(\frac{n_{13}}{n_{12}}\right)^2 r_y + \left(\frac{n_{13}}{n_{12}} - 1\right)^2 r_{12} + r_{23}$$

which is equivalent to a resistor in series with the fictitious diode V_f of $\left(\frac{n_{14}}{n_{13}}\right)^2$ times this, i.e.,

$$R_f = \left(\frac{n_{14}}{n_{13}}\right)^2 \left[\left(\frac{n_{13}}{n_{12}}\right)^2 r_y + \left(\frac{n_{13}}{n_{12}} - 1\right)^2 r_{12} + r_{23}\right]$$
$$= \left(\frac{n_{14}}{n_{12}}\right)^2 r_y + \left(\frac{n_{14}}{n_{13}}\right)^2 \left(\frac{n_{23}}{n_{12}}\right)^2 r_{12} + \left(\frac{n_{14}}{n_{13}}\right)^2 r_{23}$$

The slopes of the line above and below the zero axis are thus slightly different. Usually no great error is produced by making them the same and equal to $\left(\frac{n_{14}}{n_{12}}\right)^2 r_y$, which is useful if the winding resistances are unknown.

Equation (7) now becomes

$$L \frac{di_T}{dt} = E_s + E_b - E_4 - E_R \quad \ldots\ldots\ldots\ldots (7a)$$

where E_R is the voltage drop across the equivalent resistor.

It will be seen that if $\frac{di_T}{dt}$ is to be constant, the expression $E_s + E_b - E_4 - E_R$ must be constant. This is the horizontal distance between the curve of i_T and the line YT in Fig. 3, which is usually far from constant.

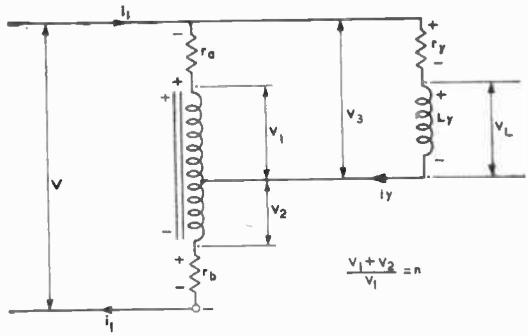


Fig. 4.—Circuit diagram including resistance of windings.

The variation of this distance is a direct measure of the non-linearity and some means of correcting for it must be found. We shall discuss this later; in the meantime let us consider various other modes of operation of the system.

4.2. *Alternative modes of operation*

In the simple mode of operation described, first the pentode takes current and then the diode. During the period A-E (Fig. 2) the diode is cut off, which means that the curve Q (Fig. 3) is adjusted to lie to the left of O, as shown at OQ′. The working point moves along the curve O U during the period A E. During E F it proceeds along the curve U V . . ., the exact shape depending on the speed of cut-off of V1. It reaches a high positive potential, returns along the zero current axis to O, and then along O Q to point X corresponding to the current $\hat{\imath}_{c2}$ of equation 1. During F D it moves from X to O. There is a region of severe non-linearity at the centre of the scan (around point O).

If the turns ratio is adjusted to move the curve of i_f to the right, such as at R Q, the non-linearity at the centre of scan is greatly reduced, since the combined curve is much smoother.

There are an infinite number of other possible modes of operation because we can control not only the position of the i_f curve but also the grid voltage of V1. For instance, if we move the i_f curve to WQ″ and adjust the grid voltage of the pentode to give the current locus NY′ the resulting curve i_T'' is that required for perfect linearity. This is not of much practical value because of the very high peak current called for from the pentode.

Reverting to the type of ampere-turn diagram used in Fig. 2 and arranging the flyback at the ends of the figure instead of in the middle gives us the diagram of Fig. 5*a* for the locus X O U of Fig. 3. Similarly, we have Fig. 5*b* for the WQ'' and NY' curves of Fig. 3. No attempt is made to depict the non-linearity in these figures. Further possible arrangements are shown at (*c*) and (*d*). Appropriate E_4-current diagrams are drawn opposite each mode. Two possible arrangements for mode (*a*) are shown, one for $V_g = 0$ for the whole cycle and one for a controlled grid voltage. Similarly mode (*d*) could be for $V_g = 0$, or fixed at some lower potential with a higher screen grid voltage. In order of goodness of linearity these modes will fall (*b*), (*c*), (*a*II), (*a*I) or (*d*), in the absence of any linearity correction. In all cases i_T, the sum of i_4 and i_f is of the same form.

4.3. *The boosted potential* ($E_s + E_b$)

We have considered the essential features of the various modes of operation; we can now reconsider some features in greater detail.

In section 4.1 it was stated that the boost potential E_b, or the charge in capacitor C2 will remain constant if the areas A B E and F G D of Fig. 2 are equal. These areas have now taken on the most diverse shapes and we have no guarantee that they will be equal. We know that the maximum excursion of i_T is fixed by the scanning requirements. The division of i_T between pentode and diode is a function of Q_r as given by equations 2 and 3, and is perhaps most easily appreciated from the diagram of Fig. 6 in which the zero line is moved for each value of Q_r to keep the peak excursion of i_T constant, but vary the contributions of positive and negative areas to satisfy equation 1. The resulting diagram is the i_T curve common to Figs. 5, and having selected a suitable mode of operation the curves of i_4 and i_f can by obtained as previously indicated. The i_f curve is a fictitious one, the actual current i_3 being larger in the ratio n_{14}/n_{13}. Our problem, then, is to so arrange this turns ratio that the integral of i_4 over the cycle is equal to that of i_3.

Writing $\frac{n_{14}}{n_{13}} = r$ we thus have the general relation that r times the area below the zero axis = area above the zero axis of the ampere-turn diagrams of Fig. 5. We notice, however, that

(*a*)

(*b*)

(*c*)

(*d*)

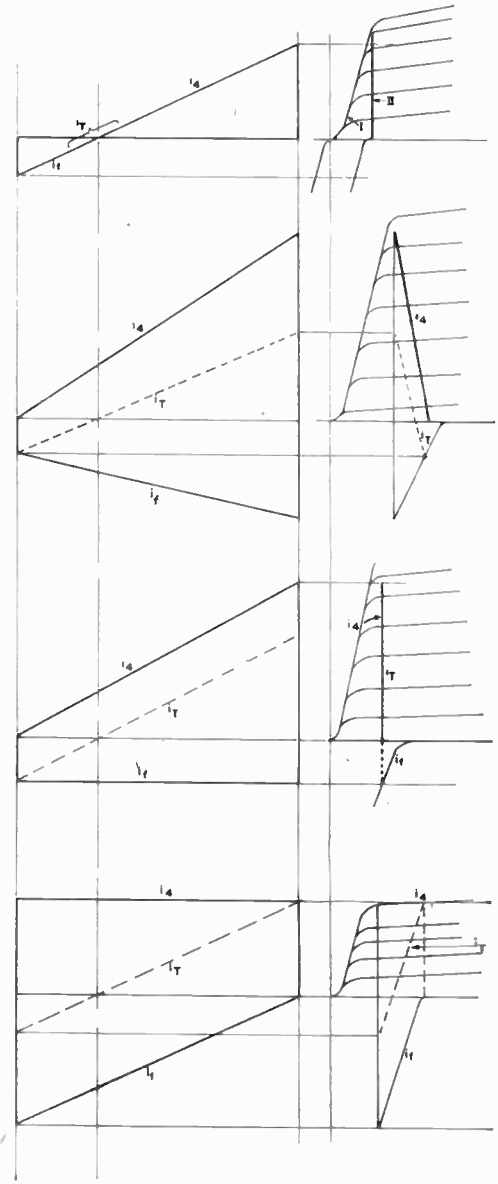


Fig. 5.—Some modes of operation, selected from an infinite number of possibilities because of their geometrical simplicity.

Mode (a) is the simplest and most efficient, (b) gives good linearity but demands high peak currents. (c) as for (b) but not so extreme. (d) is useful when the limitation is the pentode peak current; output is high but efficiency low.

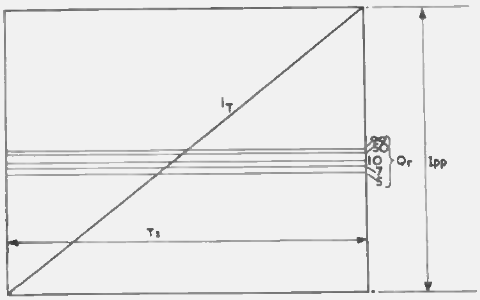


Fig. 6.—Useful for visualizing effect of Qr on the division of I_{pp} between pentode and diode.

since $i_T = i_4 + i_f$ any excess of i_4 over that required by the i_T curve immediately produces an equal excess of i_f below the axis. Let this excess be denoted by the general area A of Fig. 7, then we have

$$r = \frac{\text{positive area}}{\text{negative area}} = \frac{A + \dfrac{I_{pp}T_s}{2} \cdot \dfrac{1}{(1+\delta)^2}}{A + \dfrac{I_{pp}T_s}{2} \cdot \dfrac{\delta^2}{(1+\delta)^2}}$$

or putting $A = \beta\, I_{pp}\, T_s$

$$r = \frac{\beta + \dfrac{1}{2(1+\delta)^2}}{\beta + \dfrac{\delta^2}{2(1+\delta)^2}} = 1 + \frac{1-\delta^2}{2\beta(1+\delta)^2 + \delta^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

For certain modes, the excess area A can be written down explicitly and (8) evaluated in terms of δ. Thus for mode (*a*) we have $\beta = 0$ and (8) reduces to

$$r_a = 1/\delta^2 \quad \ldots\ldots\ldots\ldots\ldots(9)$$

For mode (c), $\beta = \dfrac{\delta}{1+\delta} - \dfrac{1}{2(1+\delta)^2}$

and we find $r_c = (1+\delta)/2\delta \quad \ldots\ldots\ldots(10)$

For mode (*d*), $\beta = \dfrac{1}{1+\delta} - \dfrac{1}{2(1+\delta)^2}$

and we find $r_d = 2/(1+\delta) \quad \ldots\ldots\ldots\ldots(11)$

For irregular curves the excess areas must be estimated and inserted in (8).

We note that once r is fixed for a particular mode the voltage E_b is determined from equation (5) which can be rewritten

$$E_b = (E_s - E_f)/(r-1) \quad \ldots\ldots\ldots\ldots\ldots(12)$$

Where E_f is the value of E_4 at which the start of the diode curve i_f has been fixed in selecting the mode, and $r = n_{14}/n_{13}$.

In some applications it is desirable to draw a steady current from C2 for operating some other portion of the set (e.g. the frame time-base) at the boosted potential ($E_b + E_s$). In this case the positive and negative areas will differ by the amount $I_b\,T$ where I_b = current drain and T is the line scanning period including flyback.

Writing $I_b = \varepsilon\, I_{pp}\, T_s/T$

equation (8) becomes

$$r = \frac{\beta + \varepsilon + \dfrac{1}{2(1+\delta)^2}}{\beta + \dfrac{\delta^2}{2(1+\delta)^2}} \quad \ldots\ldots\ldots\ldots(8a)$$

Thus the larger the drain the more r departs from unity. Equation (8*a*) can be rearranged

$$\beta = \frac{1}{r-1}\left[\varepsilon + \frac{1 - r\delta^2}{2(1+\delta)^2}\right] \quad \ldots\ldots(8b)$$

The value of E_b given in equation 12 is deduced for a condition of equilibrium; it is helpful to analyse the action of the circuit in settling down to this value after a disturbance.

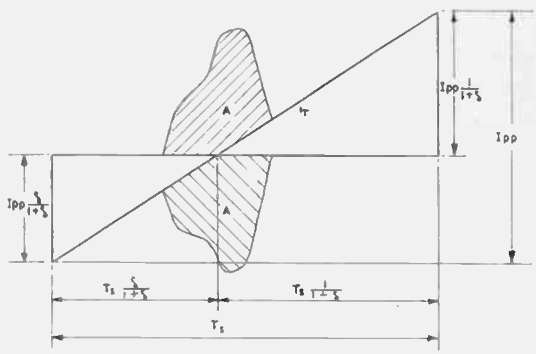


Fig. 7.—Effect of "excess area" A in producing equilibrium.

Suppose for some reason E_b falls below the correct value, e.g. by a flashover across C2. Two things will result: (*a*) point T (Fig. 3) will move to the left, and so, since ST/RS = n_{13}/n_{34} and is fixed by the transformer, point R moves to the right. (*b*) di_y/dt will fall in proportion to the drop in ST so the peak value of i_y will fall in the same ratio. The first effect (*a*) will modify the locus i_T in such a manner as to increase the excess areas of Fig. 7, i.e. to increase β in equation (8). The second effect (*b*) will decrease the peak-peak value (I_{pp}) of i_r, and so also tend to increase the value of β, which is defined in terms of I_{pp}. From equation (8) we see that the ratio *r* should now be reduced, i.e. *r* is actually higher than the equilibrium value. But the current i_3 is *r* times the area F C D in Fig. 2, so i_3 is greater than the equilibrium value. Thus i_3 charges C2 more than i_4 discharges it; the voltage E_b increases until equilibrium is again reached, and we see that the equilibrium is stable.

5. Line Time-base Generator

The problem of scanning wide angle picture tubes has focused attention on energy recovery circuits as described in the previous sections. To achieve satisfactory scanning other features of the generator must receive attention. In this section of the paper these features are discussed and finally the complete design procedure is described.

5.1. *Optimum Transformer Inductance*

Every transformer possesses leakage reactance, and this, in effect, appears in series with L_y, storing energy which should be more usefully employed. The magnetizing current of the transformer also represents storage of energy in the transformer core. A high impedance transformer increases the first loss but reduces the second, and vice versa. The maximum transfer is reached when

$$L_{12} = [k^2/(1-k^2)^{\frac{1}{2}}]\, L_y \quad \ldots\ldots\ldots\ldots(13)$$

as shown in Appendix III, where *k* is the coefficient of coupling of the transformer and L_{12} its inductance between points 1 and 2.

The upper curve $\alpha = 1$ in Fig. 8 shows the increase in VA required, over that needed by the scanning coils, for various values of *k*. The VA absorbed by the leakage inductance is plotted also, and the difference is that stored in the transformer core.

The parameter α is defined as

$$\alpha = \frac{L_{12}\text{ actual}}{L_{12}\text{ optimum,}}$$

L_{12} optimum being that given in equation (13).

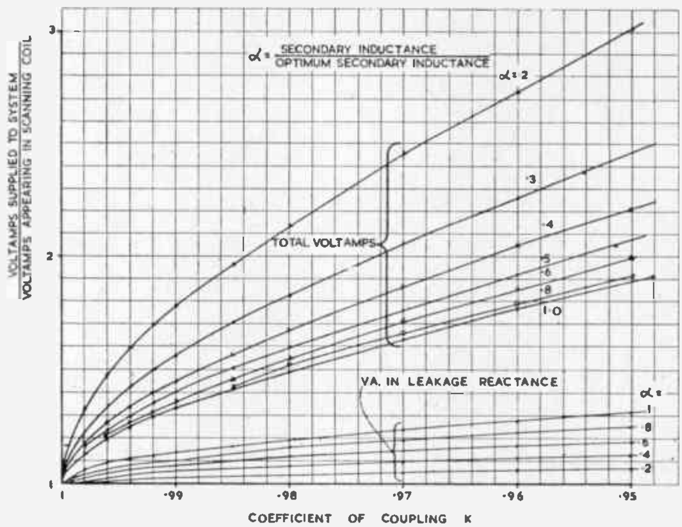


Fig. 8.—Graph showing increase in energy required due to shunt and series transformer reactances.

As α is reduced the energy in the leakage reactance falls in direct proportion, but that in the transformer core increases more rapidly as shown. Now the energy in the transformer core is largely recovered during the flyback period, but that in the leakage reactance oscillates at a higher frequency and is mainly lost. This, in itself, is a good reason for using $\alpha < 1$. A more pressing reason is the fact that these oscillations persist during the scan and cause a velocity modulation of the spot. Practical design is aimed at keeping *k* high, and α round about 0·5 to 0·8. Fig. 8 shows clearly there is no point in making $\alpha > 0{\cdot}8$. A. W. Friend[5] has produced a set of curves illustrating this point in an alternative manner.

Useful though this conception is, it is not strictly correct. A three-winding transformer has three self and three mutual inductances, and these cannot be expressed by the single variable *k*. A nearer approximation is to include a leakage inductance in each lead, making the assumption that these are not coupled to each other. Such a circuit is shown in Fig. 9 which shows a 4-winding transformer, the fourth winding, not previously considered, being used for E.H.T. production. The coefficient of coupling *k* used in equation (13) applies to transformer windings 1-4 to 1-2 for the pentode scan and 1-3 to 1-2 for the diode scan. Fortunately the condition (13) is far from critical so this distinction is mainly an academic one.

In considering the flyback, however, when the system oscillates at its several resonant frequencies, the circuit of Fig. 9 is of more value. External capacitors exist at C4, C5, C6, C7, and each winding has its own winding capacitance. An exact analysis is impossible, but useful deductions can be made.

The main transformer flux oscillates at a frequency of about 50 kc/s, and energy at this frequency is recovered. Superimposed on this are the resonances of the leakage inductances which by definition are not linked with the transformer; these we wish to suppress. The E.H.T. winding is usually built as a wave-wound "pie" to keep its self-capacitance small, and it has a large number of turns. This implies a high leakage inductance L5. The voltage change at terminal 5 during flyback is high and so the current $C_5 \frac{dV}{dt}$ in L5 is large. Thus we can deduce that the energy in L5 is greater *at the end of the flyback period* than in L4, L3 or L2.

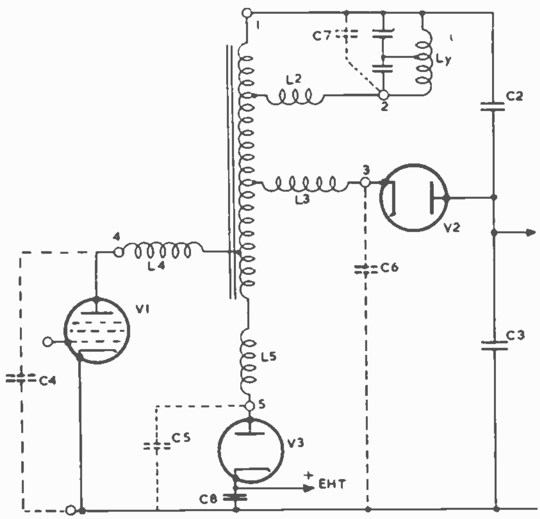


Fig. 9.—Diagram showing various leakage reactances, and external stray capacitances.

When the diode V2 conducts and holds the potential of terminal 3 constant this stored energy oscillates in the circuit of C5, L5, L3, V2, and the voltage developed across the transformer winding modulates the current in L_y. No modulating voltage could be developed across the transformer if L3 and the diode resistance were zero even though a high oscillatory current may be flowing in them. Since it is difficult to reduce L5 without increasing winding capacitance, our efforts must be turned to reducing L3. An ideal way of achieving this is to adjust the scanning coil impedance to such a value that taps 2 and 3 coincide. The diode is then connected directly to terminal 2 and this is found to reduce the effect of "ringing" very considerably. A similar effect, but of smaller magnitude occurs between C4, L4 and L3. A slightly different effect occurs at the start of flyback, when the pentode is sharply cut off. The current in L4 flows into C4 and oscillation ensues between these two components and the remainder of the system. It is possible, with an autotransformer, to make L3 and L4 very small, so this effect is not a very serious one.

We notice that C6 and L2 are not particularly troublesome. The diode current is not usually cut off rapidly; when it suddenly increases after flyback C6 is short-circuited by the diode impedance. L2 can be considered as part of L_y and enters into oscillation at the main frequency; the energy stored in it is therefore recovered.

5.2. *Linearity Correction.*

In Figs. 3 and 5*b* one method of preserving perfect linearity was shown. It is of little practical interest since it involves very high peak currents in both valves, and a carefully adjusted grid input waveform. Moreover, the power dissipation at the pentode anode is high, since it is related to the area to the left of N Y′, which, of all the loci postulated, is the furthest to the right.

Various devices have been developed to adjust linearity. Schade[2,3] has described two methods. In one the voltage applied to the diode is varied during the scanning cycle by connecting a resonant transformer between the anode circuit of V1 and the cathode circuit of V2. Energy is supplied to the transformer during the period of conduction of V1 and delivered to V2 during the remainder of the cycle, the effect being to move curve Q″ (Fig. 3) horizontally during the scan. This is shown in Fig. 10 where curves 1 to 4 show the fictitious diode characteristic at equal intervals of time during the diode conduction period, the amount of horizontal translation being adjusted to make the intercepts AB, BC, etc., of equal length. At E the diode is cut off and the pentode conduction period starts. The grid voltage is controlled to give a linear rate of rise of current from E to F.

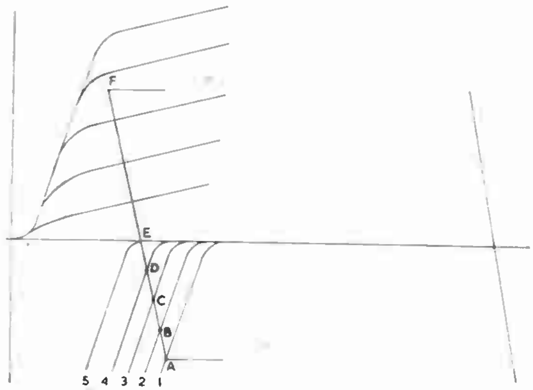


Fig. 10.—Linearization by movement of diode curve or triode equivalent.

After point E has been reached the diode curve must remain to the left of the working point. To achieve a smooth take-over between the two valves they may be allowed to conduct at the same time, and AF then becomes the difference of the two valve currents.

The difficulty with this circuit is that the pentode grid input wave must be adjusted correctly to give a linear rise of current from E to F, and at the same time this current wave must so excite the resonant transformer that the resultant movement of the diode curve gives a linear rise from A to E. By adjustment of damping and tuning of the transformer this can be achieved to a satisfactory degree, but the circuit is somewhat inflexible.

In the other method, the diode V2 is replaced by a triode. The family of diode curves in Fig. 10 now becomes the series of I_A-V_A curves of the triode. The required grid drive can be obtained from any convenient part of the circuit from which, by suitable shaping, it can be derived. There is a greater power loss in the triode than in the diode, but in the diode circuit there is a loss in the damping of the transformer Schade gives complete circuits.

We now proceed to discuss a further method which, as far as the author is aware, has not previously been described, which is based on the use of a saturated reactor. Such reactors have been used for many years as shaping elements. There was, for example, a patent granted in 1938[4] describing the use of such a reactor to "Assist the production of a saw-tooth current," but it is fairly clear from the diagrams that the application to efficiency diode circuits was not then appreciated. Saturated reactors were used in efficiency diode circuits by the author in 1946 without much success because the degree of polarization was not sufficient, but in February 1950, P. J. H. Janssen, of the Philips Co., Holland, demonstrated a circuit using such a reactor, giving a non-linearity less than 1 per cent., in which polarization was obtained from a permanent magnet.

Besides the benefit of small non-linearity, the saturated reactor is a comparatively lossless device and the energy stored in it is mainly returned to the circuit during the flyback period.

5.2.1. Design of Saturated Reactor

Figure 11(a) shows a typical B-H curve of a ferrite material such as "Ferroxcube* III" working asymmetrically between H_1 and H_2 and (b) shows a μ-H curve where $\mu = dB/dH$ for increasing values of B. Let us increase H linearly between the limits H_1 and H_2, by passing the current i_y through the winding, and providing a polarizing flux by means of a d.c. winding or a permanent magnet. Using the notation and units of Fig. 11 we can write:

$$E = N\frac{d\varphi}{dt} \cdot 10^{-8} \quad \text{(14)}$$

$$\varphi = BA \quad \text{(15)}$$

$$\mu = \frac{dB}{dH}\,;\ B \text{ increasing} \quad \text{(16)}$$

$$H = H_0 + \frac{4\pi N i_y}{10 l} \quad \text{(17)}$$

$$i_y = \gamma t - \frac{I_y}{2} \quad \text{(18)}$$

Thus

$$E = NA\frac{dB}{dt} \cdot 10^{-8}$$

$$= NA\mu\frac{dH}{dt} \cdot 10^{-8}$$

$$= \frac{4\pi N^2 A\mu}{10 l} \cdot \frac{di_y}{dt} \cdot 10^{-8}$$

$$= \frac{4\pi N^2 A\mu\gamma}{10 l} \cdot 10^{-8}$$

$$\text{or } E = \eta\mu \quad \text{(19)}$$

* Registered trademark of Mullard Ltd.

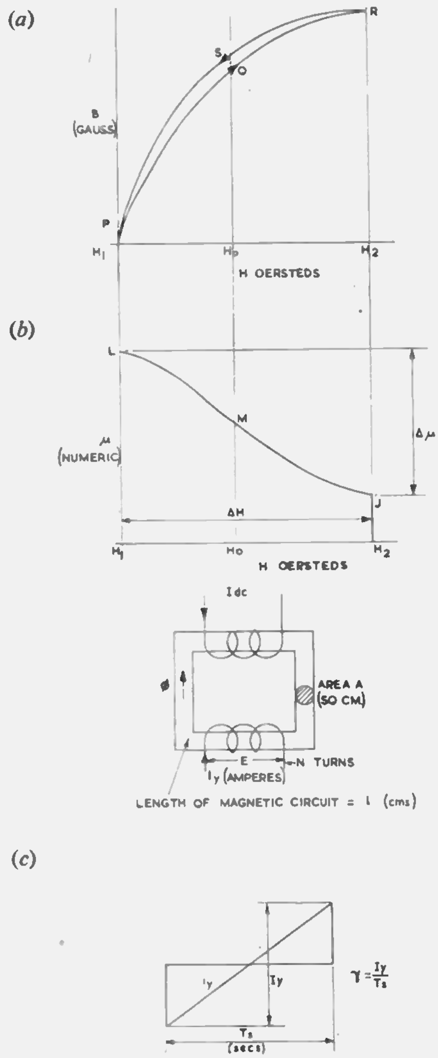


Fig. 11.—(a) B-H curve for assymetric operation. (b) μ-H curve corresponding to (a). (c) Symbols used in the analysis.

where $\eta = 1{\cdot}26 \times 10^{-8}\ N^2 A\gamma/10l$ and is constant provided di_y/dt and hence dH/dt are constants. The curve of voltage-drop versus i_y is thus of the same form as the μ-H curve. This voltage drop multiplied by the turns ratio n_{14}/n_{12} can be added to the diagram of Fig. 3 as redrawn in Fig. 12 in which the lettering is identical with those of Figs. 3 and 11(b), and the shaded area represents the votage drop absorbed by the inductor at each value of current. The intercept between the curve LMJ and the line YTZ corresponds to the (transformed) voltage across the inductive part of L_y—its resistance has been absorbed in the slope of YT. This, then, is a measure of di_y/dt and our aim is to keep it constant. Obviously, we have in our reactor a rather flexible means to achieve this end by adjustment of turns, polarizing field, core section, and possibly air gap.

The design of the reactor is a problem of curve-matching and obviously involves one initially in construction of suitable families of curves for reference.

The μ-H curve for a sample of a suitable ferrite material having a closed magnetic circuit is easily obtained oscillographically by passing the scanning current i_y through a few turns linking it, observing the voltage-drop obtained, and using equation (19) to determine the scale. Suitable families of curves can then be drawn for various values of H_0 and ΔH, and are available for curve fitting on Fig. 12. Fig. 13 shows a typical curve for H varying from 0 to 9 Oersteds, and is given as an illustration of the orders of magnitude to be expected.

Having selected a promising shape of curve let Δμ be the available range of μ and ΔH be the total excursion of H producing it.

From Fig. 12 we also know ΔE defined as n_{12}/n_{14} times the voltage difference (XL-UJ) and we know the total excursion of i_y say I_y. We can now calculate the volume of ferrite needed for a closed magnetic-circuit reactor, and from this complete the design.

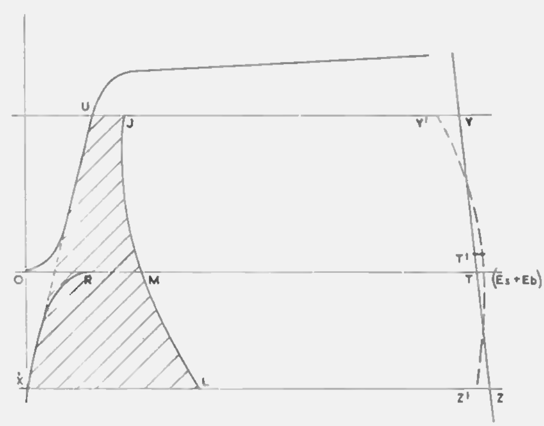


Fig. 12.—Extension of Fig. 3 to include effect of linearity reactor and capacitor C2.

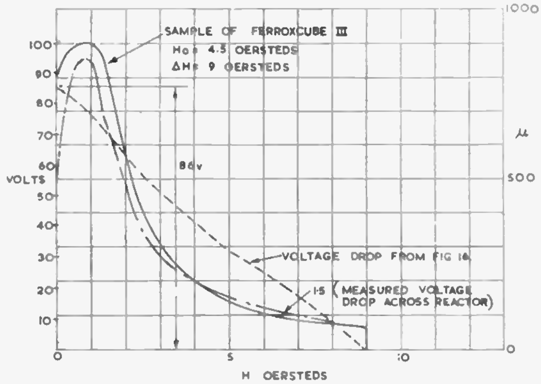


Fig. 13.—Composite diagram containing a μ-H curve from design data, an ideal voltage-time curve from Fig. 16, and a voltage-time curve measured experimentally.

For a linear rise of current, using the notation and units of Fig. 11.

$$\frac{dH}{dt} = \frac{\Delta H}{T_s}$$

Now $H = \dfrac{1{\cdot}26\, Ni_y}{l} + H_0$ Oersteds

hence $\dfrac{dH}{dt} = \left(\dfrac{1{\cdot}26 N}{l}\right)\left(\dfrac{I_y}{T_s}\right)$ Oersteds/sec.

from these we obtain

$$l = \frac{1{\cdot}26\, NI_y}{\Delta H} \quad \ldots\ldots\ldots\ldots\ldots\ldots(20)$$

Also $E = \dfrac{Nd\varphi}{dt} \,.\, 10^{-8}$

$$= \left(\frac{NAdB}{dH}\right)\left(\frac{dH}{dt}\right) . \, 10^{-8}$$

$$= \left(\frac{NA\mu\Delta H}{T_s}\right) . \, 10^{-8}\,\mathrm{V}$$

Thus $\Delta E = \left(\dfrac{NA\Delta H}{T_s}\right)\Delta\mu \,.\, 10^{-8}$ V

$$\text{or } A = \left(\frac{\Delta E\,.\,T_s}{N\,.\,\Delta H\,.\,\Delta\mu}\right) . \, 10^{8}\,\mathrm{cm}^2 \quad \ldots.(21)$$

$$\text{Volume} = Al = \frac{1{\cdot}26 \times 10^8 I_y\,.\,\Delta E\,.\,T_s}{\Delta\mu\,(\Delta H)^2}\ \mathrm{cm}^3 \quad \ldots\ldots\ldots\ldots\ldots(22)$$

This volume can be made up in any way desired. Having settled on a convenient shape, l is known, and the number of turns N is obtained from (20). If the volume is correct the area will be such as to provide the correct voltage increment ΔE across the winding, with a waveform appropriate to the curve selected.

A polarizing winding is added to produce $H_0 = \Delta H/2$. Since the volume varies inversely as $(\Delta H)^2$ the maximum value of ΔH should be used, but too high a value may produce a very non-linear curve which will be difficult to fit. The materials should always be driven well into saturation, otherwise the distance JU (Fig. 12) will be unnecessarily large and energy will be absorbed in the inductor which could have been available for scanning.

The presence of an air-gap in the magnetic circuit reduces the ratio between XL and UJ. The difference between these two and hence the mean slope can be adjusted by increasing the number of turns which increases both UJ and XL but this moves the whole curve to the right. this is again wasteful of energy which is now stored in the (linear) air-gap. In cases where this is not a disadvantage (e.g., narrow-angle C.R. tubes at moderate voltages) a very simple corrector can be made consisting of a piece of extruded ferrite tube polarized by a permanent magnet (Fig. 14).

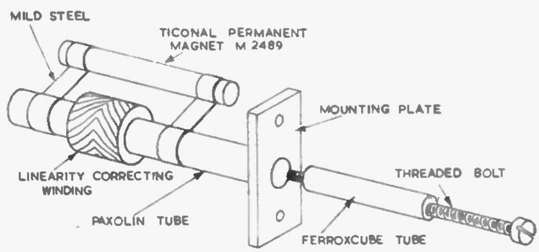


Fig. 14.—Type of linearity corrector polarized by a permanent magnet. Because of the large and variable air-gap this can only be designed empirically, but is very simple and effective in practice.

A closed magnetic circuit corrector has been described because of its theoretical simplicity; it is not without its practical difficulties. No satisfactory means have been devised of polarizing a closed magnetic circuit by means of a permanent magnet. Portions of the circuit can be polarized but it always divides itself up into two sections polarized in opposite directions which is useless for this purpose. Consequently the polarizing field is obtained from a separate winding carrying direct current.

This current must be reasonably free from a.c. and yet be fed from a source which has a high impedance to voltages appearing across the polarizing winding. It is advisable to keep the linearity corrector entirely in the line time base generator circuit and use either the pentode anode or cathode current for polarizing. Otherwise line frequency signals may be injected into other parts of the receiver and cause non-interlace.

Adjustment of linearity is conveniently made by a resistor connected across the polarizing winding. This does not vary the d.c. appreciably but acts as a positive resistor in parallel with the effective negative resistance of the reactor. It also tends to damp out self-oscillation following the flyback pulse.

A suitable design of closed-circuit reactor is described later.

An alternative way of explaining the action of the reactor is to regard it as having a negative differential resistance which cancels the positive actual resistance of the circuit. One can easily show from the preceding equations that the effective resistance dE/di_y is proportional to $d\mu/dH$ which is negative. For an idealized B-H curve, $B = aH - bH^2$, $d\mu/dH = -b$ and is constant. Actual curves depart widely from this, however, and as one must use the greatest possible range of H, this concept is only of academic interest.

5.2.2. Non-linearity due to a "Flat" Screen Face

The saturated reactor corrects for assymetric distortion. A symmetric type of distortion arises in the C.R. Tube, and can be corrected separately. If the scanning coil field is uniform

$$\sin\theta = ki_y$$

where θ is the angle of deflection from the axis and k is a constant. The movement of the spot across the screen will be proportional to θ if the screen has a centre of curvature coincident with the centre of deflection, or $\tan\theta$ if the screen is flat. These two extreme curves are drawn against i_y in Fig. 15 and actual tubes lie between them.

One way of correcting for this non-linearity is to insert a capacitor in series with L_y. The parabolic potential built up across this capacitor by the current wave, which is very nearly a saw-tooth, reduces the voltage across L_y towards each end of scan and so adjusts the velocity of the spot to a more nearly constant value.

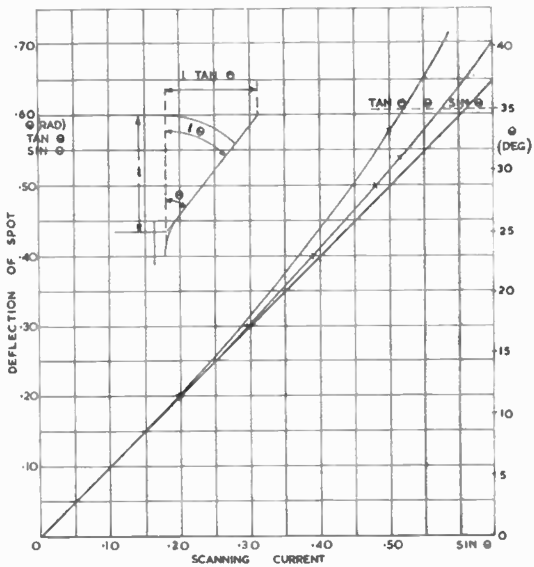


Fig. 15.—Showing non-linearity introduced by C. R. tube.

On Fig. 12 this could be represented as modulation of $(E_s + E_b)$ so that the line YTZ takes up the position Y′T′Z′ (T′ is equidistant from lines JY and LZ).

The finite value of C2 (Fig. 1) will produce a similar effect by actual modulation of E_b. The curve is tangential at T instead of T′ and the two half-parabolæ have different laws corresponding to the unequal figures ABE and FGD of Fig. 2, but the difference is slight. Suitable choice of C2 thus provides an economical way of achieving the same end as an additional capacitor in series with L_y.

It is apparent that curvature of the reactor characteristic such as that shown by LMJ in Fig. 10 will produce the same effect as C2 and in practice both are adjusted by experiment for the best compromise.

5.3. *Width Control*

There are at least four types of width control available.

(*a*) A variable reactor in series with L_y.

(*b*) A variable reactor in parallel with a portion of the transformer.

(*c*) A combination of (*a*) and (*b*).

(*d*) Variation of the h.t. supply.

(*a*) is obviously equivalent to varying L_y. (*b*) amounts to the same thing because it can be regarded as a (transformed) inductor in parallel with L_y. (*c*) can be adjusted so that the impedance of the combination does not change appreciably, and then it is equivalent to a change only in the sensitivity of L_y. This is obviously better than (*a*) or (*b*) alone because there is less reaction on the circuit. Because the inductances cannot be made zero or infinity, however, there is always a fixed absorption of energy. To use (*d*) one must design for a lower h.t. voltage than is available. However, when circumstances require maximum scan the full voltage is available; a resistor can be reduced to zero. The disadvantage is a severe reaction on the circuit, particularly E.H.T. supply and V_b. By supplying the frame circuit from the boosted potential the frame scan can be made to vary with the line scan and E.H.T. The picture then merely changes in brightness, the change in size or aspect ratio being slight.

5.4. *E.H.T. Supply*

If L_6 is the inductance of the system between terminals 1 and 4, C its equivalent capacitance at the main flyback frequency, and I is the value of i_r (Fig. 5) at the instant the pentode is cut off, then at the instant of the voltage peak V_4 at terminal 4 (Fig. 9) we can write

$$\tfrac{1}{2}L_6 I^2 = \tfrac{1}{2}CV_4{}^2 \text{ for a lossless circuit}$$

$$\text{or } V_4 = I\sqrt{\frac{L_6}{C}}$$

Taking Q_r into account this becomes

$$V_4 = e^{-0\cdot 83/Q_r} I\sqrt{\frac{L_6}{C}} \quad \ldots\ldots\ldots\ldots(23)$$

C is the equivalent capacitance

$$C = C_4 + \left(\frac{n_{15}}{n_{14}}\right)^2 C_5 + \left(\frac{n_{13}}{n_{14}}\right)^2 C_6 + \left(\frac{n_{12}}{n_{14}}\right)^2 C_7$$

plus the effect of winding capacity.

This voltage is stepped up by the overwind 4-5, and rectified by V3. Too high a ratio increases L_5 and ringing, and lengthens the flyback time. Moreover L_5 is in series with V3 and tends to restrict the current pulses charging C8 so causing poor regulation. In practice excess wire is wound on and gradually stripped off until the best compromise is reached between these factors.

It does not seem to be generally appreciated that the E.H.T. regulation obtainable from energy recovery scanning circuits is fundamentally poorer than that obtained if the scanning energy is dissipated and supplied afresh during each cycle of operation. The reason why this may not be so in practice is a rather incidental one.

Consider an idealized system in which energy W is supplied during the scan, stored in a capacitance C during the flyback, and finally dissipated in a resistor. We write:

$W = \frac{1}{2}CV^2$ where $V =$ unloaded E.H.T. voltage. Let a load current I produce a voltage drop v. Then E.H.T. energy drawn per cycle $= IT(V-v)$ where $T =$ period of one cycle. This must equal the difference between W and the energy remaining in the condenser or

$$IT(V - v) = W - \tfrac{1}{2}C(V - v)^2$$
$$= \tfrac{1}{2}C[(V^2 - (V - v)^2] \,.$$

from which

$$\text{Source Impedance} = \frac{v}{I} = T/C \quad \ldots\ldots\ldots(24)$$

Now consider another idealized system in which the losses are zero; energy W is supplied when the set is switched on and proceeds to scan the tube indefinitely. Any drain of power, however small, would eventually reduce scan and E.H.T. to zero. The source impedance $\frac{v}{I}$ is thus infinity.

In so far as actual energy recovery systems lie between these two extremes the source impedance may be expected to lie somewhere between T/C and ∞. This is easily understood since an E.H.T. drain reduces the effective Q during flyback and thus the energy returned to C2 (Fig. 1) is reduced. The voltage available for the next scan is therefore less and so the current amplitude is smaller. The recovered energy is then still less, and so the cycle goes on until equilibrium is reached at a lower E.H.T. voltage. If we assume that the mean power P supplied to the inductive circuit does not change with E.H.T. current it is not difficult to show that the initial source impedance of the timebase is approximately $V^2/2P$. Obviously the more efficient the circuit the worse the regulation will be.

In an actual circuit for 625 lines, the mean power delivered to the circuit was 11 W and the

E.H.T. voltage was 12 kV. The calculated source impedance is 6·5 MΩ and the measured value was 7 MΩ.

If the incoming power increases with increasing E.H.T. drain a better regulation will be obtained. When the mode of operation employs the pentode in the knee of its $I_A V_A$ curve this can occur by the following action. The E.H.T. drain reduces the recovered energy and the potential E_b falls. This, as shown earlier, causes the diode curve of Fig. 3 to move to the right. Now if the pentode is operating to the right of the knee its current is controlled by its grid voltage and the change in its anode potential has little effect. Consequently the diode current waveform is also mainly unchanged and hence the power supplied is unchanged. If, however, the pentode is operating in the steep portion to the left of the knee the change in anode potential has a marked effect on the current it takes, its conduction period commencing earlier as the diode moves to the right. This, in turn, means that the diode current curve falls away more slowly (vide equation (6)) and hence the area under it is larger. But the area under this curve multiplied by the voltage E_s is the power supplied to the system, which has, therefore, increased.

Other reasons can be given for working the valve in this region. The lower part of the characteristic varies very little with screen voltage or with valve ageing. The screen dissipation is high, however, and must be carefully checked.

In the example cited, the good agreement between 6·5 and 7 MΩ is rather fortuitous. The action just described reduces the source impedance by about 20 per cent., while the leakage reactance L_5 and E.H.T. diode impedance increases it.

The heater supply for V3 is obtained from the energy in the transformer by winding on a few turns of suitably insulated wire for the purpose. The power loss affects Q_r and therefore δ.

5.5. *Supply for Heater of Efficiency Diode*

The heater supply for V2 is obtained from the normal heater line, or from a heater transformer designed for the purpose.

The d.c. potential of the cathode will be $E_s + E_b$ above earth. Superimposed on this will be the pulse voltage developed across winding 1-3. The heater potential with respect to earth depends on the type of heater supply. For parallel connected heaters it is substantially at earth potential. For series-connected heaters it will be at an a.c. or d.c. potential above earth corresponding to the point in the heater chain at which it is connected. Using a low-capacity transformer, it is possible to tie the heater winding to the cathode, but this is expensive besides being impracticable with d.c. mains supply.

The transformer can be avoided by supplying the heater current via a bifilar winding equal in turns to n_{13} and wound on the same core. This applies to the heater the same pulse voltages as those on the cathode. The d.c. voltage difference $E_s + E_b$, plus the supply-frequency voltage, plus the line frequency ripple on C2 remain.

The valve-maker sets a limit to the permissible d.c. voltages between heater and cathode and this limits the permissible value of E_b when a bifilar heater supply is used.

The bifilar winding is usually wound on the opposite leg of the transformer from the main winding because a high k is not so important. Nevertheless care must be taken with leakage inductance ringing since the oscillations are applied between heater and cathode and may cause the limit to be exceeded. A capacitor of 0·02μF between cathode and heater prevents this from becoming serious. Too high a value of capacitance will inject supply frequency voltages into the timebase causing distortion.

5.6. *Drive Circuits*

A grid voltage waveform, appropriate to the selected current and anode-voltage waveforms, must be applied to the pentode. By suitable passive shaping circuits these can usually be derived from one of the transformer tapping points, or special ratios may be provided. In such a case a single-valve time-base is obtained. Many have been described, each with its merits and demerits. In many cases the unsynchronized repetition frequency is controlled by the working-point reaching the "knee" of the $I_A - V_A$ curve. This is unsatisfactory because as the valve ages the current corresponding to this point falls, and the repetition frequency rises. Eventually the free-running frequency may exceed the synchronizing frequency, and the time-base cannot be locked. In other cases the bend of the B-H characteristic has an effect and as this varies with temperature it is even less satisfactory.

The drive circuit employed in the circuit given later (Fig. 19) uses the triode portion of an ECL80 triode-pentode valve to phase-reverse and clip a voltage obtained from a tap on winding 1-2. This has several advantages. By selection of the time constant in the triode anode various degrees of curvature of the wave-front can be obtained giving control over the pentode current locus. In particular, this enables screen-grid dissipation to be minimized. By clipping the transformer voltage, interaction between E.H.T. drain and drive voltage is prevented. The repetition rate is determined by the R-C network in the triode grid circuit, and is independent of valve ageing or temperature. Unlike the conventional blocking-oscillator no additional transformer is required.

5.7. *Example of Design Procedure*

5.7.1. Theoretical Steps.

Let us take as a typical problem the design of a timebase to scan at 405 lines a 16-in. metal tube MW41-1 using a "Double-D" picture format. The E.H.T. is to be 14 kV, rectified by a single EY51 valve. The valve operating limits must be observed and of these the peak inverse voltage of the diode and, if a bifilar heater winding is used, the heater-cathode d.c. voltage of the diode must be considered early in the design. With a bifilar heater winding the d.c. voltage between heater and cathode of the efficiency diode is equal to the boosted voltage ($E_b + E_s$). This design was carried out during the development stage of the PY80 efficiency diode and the preliminary limit of approximately 450 V has been used. The H.T. line is 190 V, as the set is intended for d.c./a.c. operation at supply voltages down to 200 V. Non-linearity is to be less than 5 per cent. A drain of 2·8 mA at 450 V is required for the frame timebase oscillator.

Figure 16 shows the curves of Fig. 3 drawn to scale. Valve V1 is a PL81, and in the upper part of the diagram is drawn the envelope of the $I_A - V_A$ curves of this valve. For sake of safety, in case of failure of drive, a cathode bias of 25 V is included, so moving the curve 25 V to the right.

The voltage $E_s + E_b$ is fixed at 450 maximum, so point T is at 450, and S is at the H.T. voltage of 190.

We have seen that it is advantageous to have the diode and scanning coil taps coincident. The value of scanning coil inductance has, therefore, been adjusted, by selection of wire gauges, to make the voltage-drop across coils, linearity corrector, and width control equal to the boost voltage E_b. Appendix II gives details of their construction. The width control is of the series-parallel type which maintains the effective inductance presented to the transformer approximately constant. It decreases the sensitivity by 6 per cent. at the maximum width adjustment and 24 per cent. at minimum width setting, so we can take the mean drop in sensitivity at the centre of adjustment as 15 per cent., with available variation of $\pm$ 10 per cent. about the centre value.

The scanning coil current for full scan is 545 mA peak-peak at 14 kV on the MW 41-1 C.R. tube. We thus require 545 × 1·15 or 630 mA peak-peak. The inductance is 32·3 mH and the voltage drop is thus 240 V, leaving 20 V for the linearity corrector at the mean scan conditions.

From previous experience we know we can achieve $k = 0{\cdot}998$.

Thus

$$\frac{L_{12}}{L_y} = \frac{k^2}{\sqrt{1-k^2}} = \frac{\cdot 996}{\sqrt{\cdot 004}} = 16.$$

At these high values of k there is little loss with quite low values of α so we take $\alpha = 0{\cdot}5$ and $L_{12}/L_y = 8$. The increase in energy required is found from Fig. 14 to be 16 per cent.

Thus 1·16 × 0·63 × 240 or 176 VA is equal to the peak-peak value of i_r multiplied by the voltage MT of Fig. 16. Now MT is not yet known accurately, but can be estimated within 10 per cent. or so which is near enough at this stage. Taking MT = 400 V gives $I_{pp} = 440$ mA. Of this the pentode provides slightly more than half, say, 240 mA. It is well to select a mode in which the diode does not quite cut off at the end of scan, and for reasons previously given we like to work the pentode below the "knee." Thus the root of the diode curve will be immediately below, say, the 240 mA point of the pentode curve, i.e., at + 60 V. This fixes point R within about 5 V and gives us our turns ratio $r = RT/ST = (450 - 60)/(450 - 190) = 1{\cdot}5$ within 1 per cent.

We can now find the slope of the line YT fairly accurately as $(1{\cdot}5)^2$ ($Ry + R$ width) or $2{\cdot}25\,(29{\cdot}1 + 11) = 90\,\Omega$. Later we will adjust this for resistance of the linearity control and transformer. At this stage the current swing and

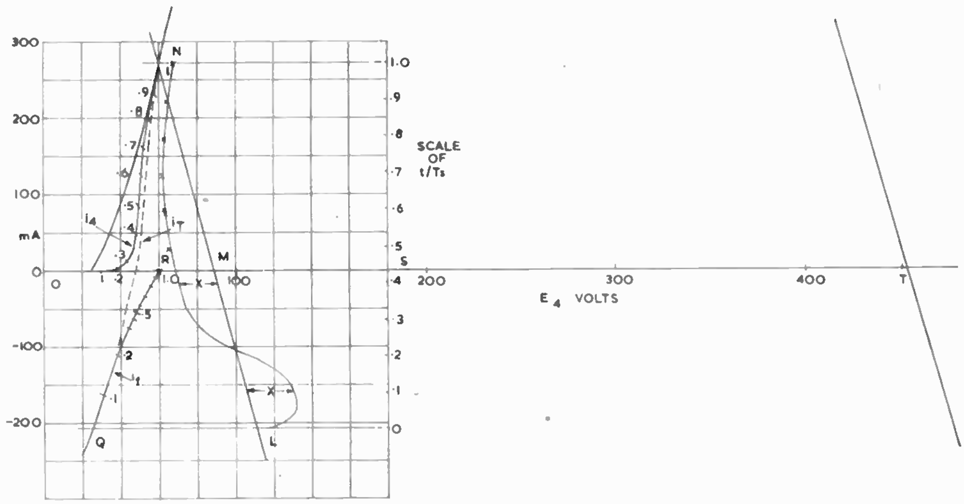


Fig. 16.—Current-voltage diagram for a typical design.

the peak current in the pentode should be checked to confirm or adjust the first estimate.

Now in order that E_b, fixed arbitrarily, shall be the equilibrium value, we must give β in equation 8(a) the value required for $r = 1{\cdot}5$ by selection of the appropriate mode. This involves a knowledge of δ, i.e., Q_r. The factors affecting Q_r are the losses in the deflector coils, transformer, and E.H.T. rectifier heater, the E.H.T. current drain and the possibility of V1 not being cut off quickly enough at the start of flyback. Schade[3] gives details for calculating some of these but in practice δ can often be estimated more accurately from previous designs in which the peak diode current and peak pentode currents have been measured.

We shall take $\delta = 0{\cdot}76$ from previous work and thus find the peak pentode current to be 272 mA, and peak diode current 208 mA. Thus we are able to draw the line NM through $I_4 = 272$ mA parallel to YT as the idealized locus about which our curve-fitting must be adjusted.

Since the highest ratio obtainable is that of mode (a) Fig. 5, in which $r = 1/\delta^2$ we check $r_{max} = 1/\delta^2 = 1{\cdot}73$ so it is possible to obtain $r = 1{\cdot}5$ with some "excess area," i.e., β is positive.

The boost drain factor ε is found to be $\cdot 0058$, and thus from equation 8(b) $\beta = 0{\cdot}0536$. Our mode must be such as to obtain this, but we need not worry about it too much because the system will automatically adjust itself to the correct value. All we need do at this stage is to see that there is some overlap between the diode and pentode curves of Fig. 16 and wind our experimental transformer with a few taps on either side of the calculated ratio for n_{14}. It is preferable to arrange n_{12} so that the tap 2 comes at the end of a layer, and make adjustments of ratio by n_{14}. The slight variation in L_y/L_{12} is of no consequence. We then choose the tap which gives the correct value of E_b, when $\beta = \cdot 0536$ will be automatically achieved. Alternatively, one can obtain a graphical solution, but the experimental method is very rapid. In Fig. 17 both methods are shown. We plot I_{pp} vertically and find the area A which is $0{\cdot}0536$ times the total area. Triangular figures of area A (diode) and $A - \varepsilon T_s$ (pentode) are added as shown, so arranged that the diode is just conducting at the end of scan. The actual i_4 and $i_3.n_{13}/n_{14}$ curves of the finalized design are given for comparison. The agreement of the excess areas is well within experimental error. The actual current waveforms will, of course, not, in general, be straight lines in the overlap area but the difference curve i_T should be. The difference curve lies within 3 mA of the theoretical straight line at all points.

Proceeding with the theoretical design we draw a vertical time scale t/T_s on Fig. 16 and proceed to plot out the locus of i_T i_4 and i_f from the information in Fig. 17. We plot first i_f, making a time scale along the diode curve. At the resulting values of E_4 we plot the pentode currents, and i_T locus. The curve i_T must lie such that $i_T = i_4 + i_f$ because we made it so in Fig. 17. We could now insert the grid voltage contours for V1 and determine the grid-input wave required, but that would be merely pedantic. There is no great virtue in our straight lines of Fig. 17. The main point of this is to find the i_T contour, and so find the area on Fig. 16 to be absorbed by the linearity corrector. The dotted curve in Fig. 13 has the same shape as the voltage difference between LN and the i_T curve of Fig. 16 but to a scale which makes it the best fit (by eye) to the μ-H curve. Based on the incremental permeability curve of Fig. 13 the voltage across the linearity control is drawn in on Fig. 16. By choosing a smaller value of ΔH a better fit could be obtained, but this increases the volume of Ferroxcube required and a really accurate fit is not essential as we shall see.

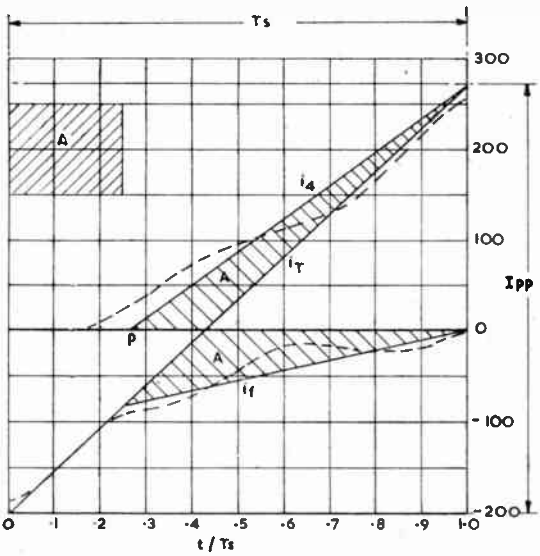


Fig. 17.—Full lines—calculated current diagrams. Dashed lines—measured current waveforms.

Equation (22) is not altered by writing $1{\cdot}5\ \Delta E$ instead of ΔE, and $I_y/1{\cdot}5$ instead of I_y. Now $1{\cdot}5\ \Delta E$ is the voltage drop QL of Fig. 16.

Thus $T_s = 85 \times 10^{-6}$ secs.

$\Delta H = 9$ Oersteds .

$1{\cdot}5\ \Delta E/\Delta\mu$ from Figs. 13 and 16 is 86/770 or 0·111 V.

$I_y/1{\cdot}5 = 0{\cdot}36$ A. (This is less than I_{pp} because the magnetizing current and current in parallel width control do not flow in the reactor.)

From equation (22) the volume of Ferroxcube required is 5·3 c.c. We find a pressing of the section shown in Fig. 18 which, using two pieces, amounts to 6·5 c.c. The discrepancy can be resolved by adjustment of turns since small variations in ΔH will not affect the general shape

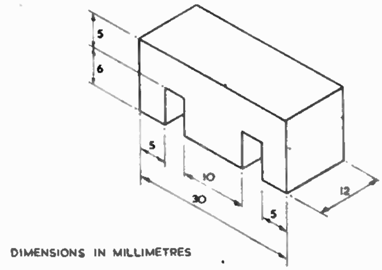


Fig. 18.—"Ferroxcube" core for linearity corrector.

of the curves. As the actual volume is 22 per cent. too large, ΔH needs to be adjusted to 9 per cent. less, or 8·2 Oersteds. The magnetic length is 6 cm, allowing for two unavoidable air gaps of ·0005 in., and $\Delta H/I_y = {\cdot}15$. From equation 20 we find $N = 0{\cdot}8l\ \Delta H/I_y = 72$ turns. We wish to adjust linearity, so we wind on about 40 per cent. excess, giving 100 turns. The polarizing winding must be adjusted, according to the current supply available, to give $H_0 = 8{\cdot}2/2$ or 4·1 Oersteds.

Plotting the μ-curve of Fig. 13 according to the voltage scale $\Delta E/\Delta\mu = 0{\cdot}111$ on Fig. 16, enables us to calculate the "current" non-linearity at any position of the scan. The maximum non-linearity occurs at X, and is of value 20/365 or 5·4 per cent. This is, in fact, quite satisfactory, because the curvature is in the right sense to correct for symmetric distortion produced in the C.R. tube. Suitable adjustment of C2 will, therefore, make the linearity on the tube face well below the 5 per cent. limit.

In the actual circuit the voltage across the linearity reactor can be checked with an oscilloscope against the expected curve, as shown in Fig. 13.

The final theoretical step is to calculate the core size and winding details to give the correct inductances, resistances and current handling capacities. This is familiar ground, but the flux density in the core should be carefully checked to avoid trouble in the experimental stages.

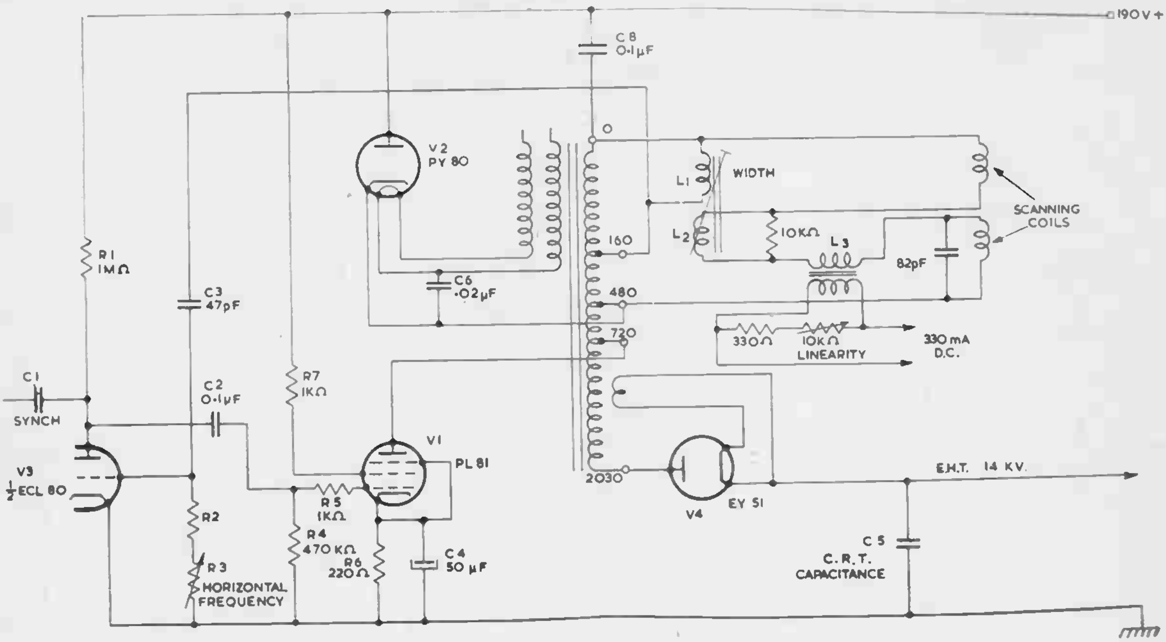


Fig. 19.—Circuit diagram of line time-base.

The a.c. component of the total flux is simply given by $\varphi = E.T_s/N \times 10^8$ lines where E is the voltage across any N turns during scan. The steady component is produced by the mean anode current flowing through n_{34} and the flux is calculated in the usual way by finding H and drawing a "gap-line" corresponding to the air gap reluctance through this value of H to intersect the B-H curve of the material at the flux density required. The total flux density must not cause the core to saturate, as otherwise the magnetizing current may rise rapidly at the end of scan, so causing the diode to cut off, and produce cramping. Because the magnetizing current is usually a small proportion of the total this effect is not often serious. With ferrite materials, however, there is a marked fall in the saturation flux density with increased temperature so any tests for this effect must be done with the transformer hot.

5.7.2. *Practical Stage*

At this point the design is fairly complete on paper and passes into the laboratory stage. We have considered the case where E_b is limited. If no limit is placed on E_b the design proceeds to minimize input power, allowing E_b to vary as required. There may be other limitations, such as peak currents, and one may be forced to use a mode similar to Fig. 5*d* in order to use the pentode at its maximum current rating over the whole period.

The diagrams and conceptions described, however, are equally applicable to any of these problems. Similarly, the laboratory procedure given below forms a basis for the practical design of any efficiency diode circuit.

(1) Wind the transformer, without the E.H.T. winding, bringing out additional taps on either side of those required for the theoretical ratios.

(2) Assemble transformer with bifilar winding.

(3) Measure winding inductances coupling coefficient and resistances. Check L_{12}/L_y and α.

(4) Operate transformer in the circuit adjusting R3 and C3 (Fig. 19) for correct frequency.

The drive wave form varied by the resistor R1 should be adjusted to give the best ratio of scan to H.T. current at this stage.

Examine a plain raster for velocity modulation i.e., leakage inductance rings. The grid and cathode of the picture tube must be shorted to a.c. signals for this test.

The capacitor C7 across one half of the deflector coil must be adjusted for minimum ringing.

If rings still are apparent the gauge of the wire and the winding arrangement must be adjusted to minimize the leakage inductance.

(5) Connect linearity and width controls

Damp the series width coil L2 with components R10 and C9 suitably adjusted. Check operation of width control. Check linearity including effect of capacitor C8.

(6) Measure oscillograms and d.c. values of PL81 cathode current, screen current, PY80 anode current, PL81 screen voltage, and boosted voltage. Measure the peak voltages on the PY81 anode and PY80 cathode. It should be remembered that these will be reduced by approximately 20 per cent. when the final transformer is assembled and waxed.

(7) Check the mode of operation and the efficiency of the circuit and adjust the ratios as required employing the extra taps.

(8) Repeat operations (4) (5) and (6).

(9) Wind a further transformer employing the experimentally checked ratios but still without the E.H.T. overwind and check as before. If the transformer is layer wound the windings should be arranged so that the taps occur at the end of a layer.

(10) Wind on the E.H.T. overwind calculating the number of turns required by direct ratio from the previously observed peak voltages and adding an allowance of 30 per cent. to cover increased capacitance and to allow for some adjustment.

(11) Wind EY51 heater winding and check the filament temperature by comparison and substitution with a normally a.c. heated EY51.

(12) Connect EY51 in circuit and check E.H.T. and adjust turns so that the E.H.T. is approximately 5 per cent. high to allow for waxing.

(13) Check for rings and flyback time. The winding arrangement may be adjusted as a compromise between ringing and flyback time. If the required E.H.T. cannot be obtained, then the transformer inductances must be reduced so that more energy is stored in the circuit at the end of the scan. This step will involve a complete redesign of the circuit.

(14) Check E.H.T. regulation, peak voltages on valves and all other valve limits.

(15) Wax impregnate in the normal manner and finally apply a thick smooth layer on the outside edge of the E.H.T. overwind.

(16) Recheck E.H.T., E.H.T. regulation, peak voltages and other operating conditions.

(17) Adjust C6 so that no leakage inductance rings exist between heater and cathode of the PY80, V2.

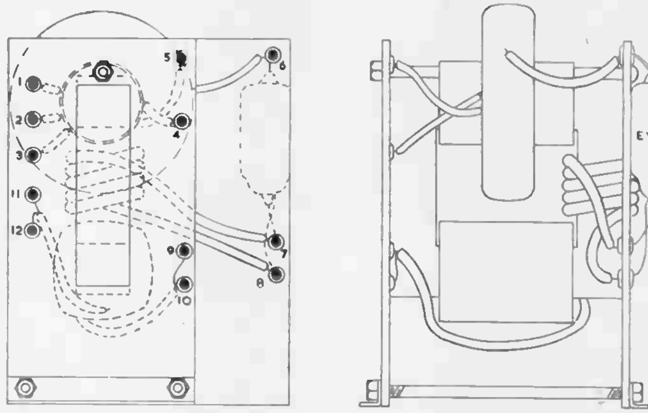


Fig. 20.—Assembly of line transformer.

(18) Carry out an overvoltage test by increasing the H.T. until the E.H.T. is increased by 50 per cent. Examine the transformer for corona in a dark room.

A complete circuit of the line timebase generator which has been taken as an example is given in Fig. 19 and the transformer is shown in Fig. 20. The scanning coil is described in Appendix II.

The procedure outlined above has been carried out and all currents and voltages were found to be within the valve ratings and close to the calculated values. The non-linearity was less than 5 per cent. over 95 per cent. of the picture.

There was a region of 5 per cent. of the picture at the left-hand edge where non-linearity reached 20 per cent. This would probably be acceptable in practice, especially as a "double D" format

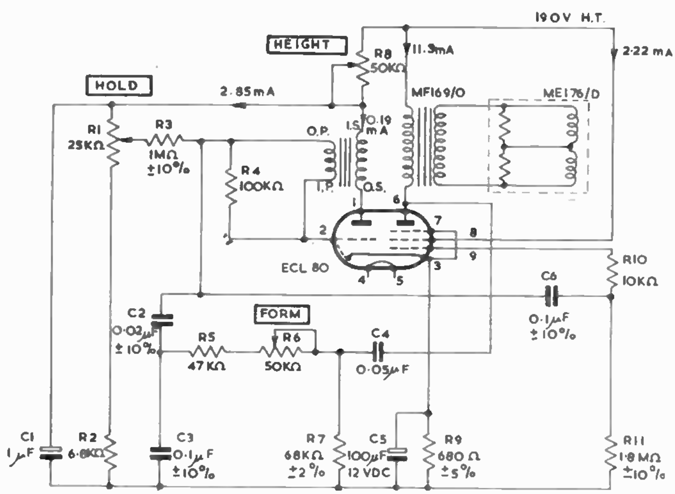


Fig. 21.—Circuit diagram of frame time-base.

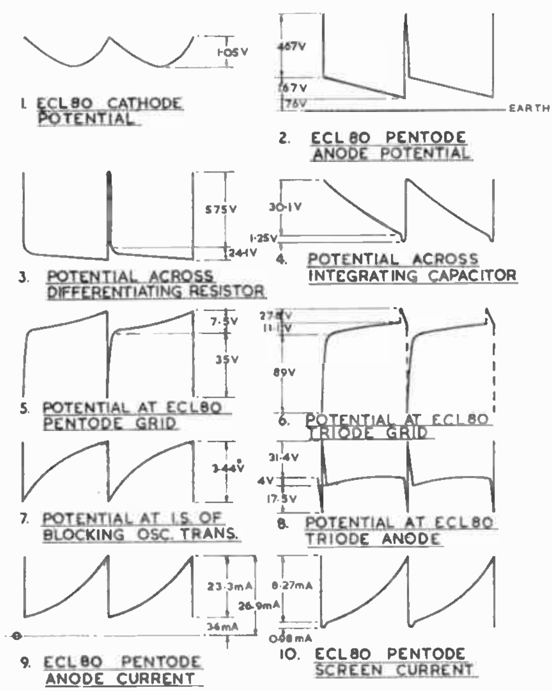


Fig. 22.—Frame time-base waveforms.

has been assumed so that most of the edge is lost, but it would be desirable to eliminate it if possible.

The non-linearity is caused by the oscillation of the scanning-coils with their associated capacities persisting into the scan period. Thus, at the commencement of scan, the coils are still receiving current from the capacities and the full scan current does not, therefore, appear in the linearity corrector. The correction voltage, therefore, departs from the calculated value, as shown by the curves of Fig. 13, in which the departure is greatest at the commencement of scan. If the actual voltage is transferred on to Fig. 16 the cause of the non-linearity will be immediately apparent.

It was found that the magnitude of this phenomenon is greatly reduced if the width and linearity coils, L2 and L3, are connected between the two halves of the scanning coils as shown in Fig. 19.

6. Frame Timebase Generator

The introduction of wide-angle scanning poses no new problem in frame scanning; normal valves can supply the required power. For the sake of completeness a frame timebase generator circuit is given in Fig. 21 for use with the deflector coils described in Appendix II and an E.H.T. of 14 kV. The circuit employs a small triode-pentode, ECL80, and the well-known feedback linearity circuit due to Blumlein. Various waveforms are given in Fig. 22.

7. Conclusions

Wide-angle scanning makes great demands on the line timebase generator. Recently introduced techniques, materials, and valves described in this paper, allow scanning and E.H.T. circuits to be designed with relatively better performance than circuits for narrow-angle tubes at present used in commercial receivers.

8. Acknowledgments

Acknowledgments are due to Mullard, Ltd., for permission to publish this paper and in particular to Dr. C. F. Bareford, manager of the Mullard Research Laboratory, who afforded every conceivable help and encouragement.

It is impossible to acknowledge in detail the source of all the material herein presented. The analysis is the author's, but the underlying ideas have been shaped by daily contact with many other workers of whom Messrs. C. H. Braybrook, S. N. Doherty, E. T. Emms, K. E. Martin and B. R. Overton are those mostly concerned with this work.

9. References

1. A. D. Blumlein, British Patent 400,976 (application date: April 4th, 1932).
2. O. H. Schade. "Magnetic Deflection Circuits for Cathode Ray Tubes," *R.C.A. Review*, **8**, Sept. 1947, No. 3, pp. 506-538.
3. O. H. Schade. "Characteristics of High-Efficiency Deflection and High-Voltage Supply Systems for Kinescopes," *R.C.A. Review*, **11**, March 1950, No.1 pp. 5-37.
4. British Patent 493,142 (Ferranti, Ltd., and M. K. Taylor. Feb. 24th, 1938).
5. A. W. Friend. "Television Deflection Circuits," *R.C.A. Review*, **8**, 1947, p.122 (Fig. 4).

Appendix I: Efficiency Diode Circuits

In the body of the paper the auto-transformer circuit of Fig. 23.2 has been described. Three other circuits are shown in Fig. 23 and in this appendix their features, advantages and disadvantages are briefly listed.

(1) *Efficiency Diode in Primary; Diode connected to Negative Supply Line* (*Fig.* 23.1)

The primary of the transformer is across the H.T. line when the diode is conducting. The overwind for the diode is arranged so that the pentode anode is at the knee potential.

Advantages:

1. Auto-transformer connection can be employed.
2. No boost potential is produced. (The boost voltage can be an embarrassment when the H.T. line has already a high value.)

Disadvantages:

1. Cathode of the efficiency diode is subject to very large peak voltages. In practice, the heater has to be supplied from a winding on the transformer.
2. The voltage across the primary is not increased in this circuit and the peak and mean currents through the valves are high. The requirements for the diode are thus particularly severe—low heater wattage, low anode impedance, and high peak inverse voltage.
3. D.C. flows through winding of transformer so capacitor is needed to prevent d.c. in deflector coils.
4. No boost potential available for rest of receiver.

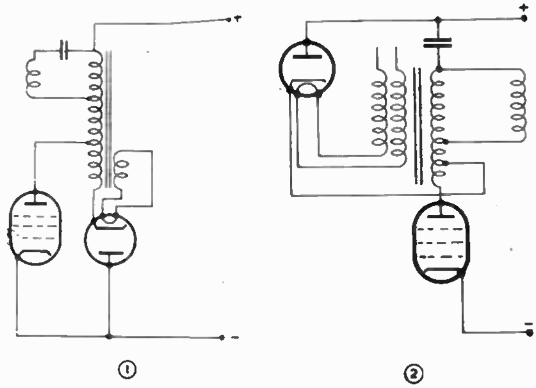


Fig. 23.—*Alternative forms of energy-recovery scanning circuits.*

(2) *Efficiency Diode in Primary; Diode connected to Positive Supply Line* (*Fig.* 23.2)

The efficiency diode is connected between H.T. line and a tap on the primary of transformer. Heaters of the efficiency diode are fed through a bifilar winding so that heater has the same peak a.c. potential as the cathode.

Advantages:

1. Auto-transformer connection can be employed.
   Unless other parts of the receiver are fed from the boosted potential no d.c. flows in the diode portion of the primary. Hence no capacitor is required in series with the coils.
2. A boost potential is available for other parts of the receiver, a useful feature for receivers with a low value H.T. line.
3. Valves are in series and the mean current through them is low. This advantage applies to all the circuits except (1).

Disadvantages:

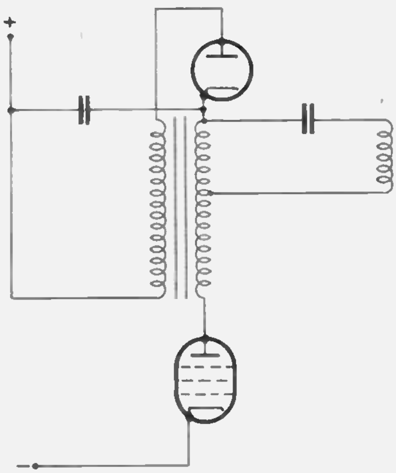
1. Bifilar winding for diode heaters takes a great deal of space on the transformer.

2. The steady potential between heater and cathode of the efficiency diode is equal to the boosted potential.

   This may prove to be a limitation in design.

(3) *Efficiency Diode in Primary with additional Tertiary winding; Diode connected to Positive Supply Line.* (*Fig.* 23.3)

A further winding is connected between the diode and the H.T. line such that the diode anode is positive during the scan and negative during the flyback.



③

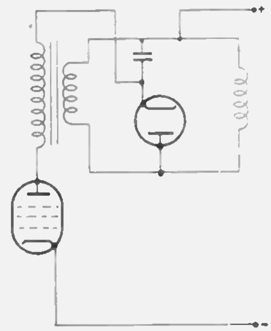
Fig. 23.

Advantages:

1. Auto-transformer connection is possible. (In each case the secondary is considered to be the winding supplying the deflector coils though in Figs. 23.1, 23.2 and 23.3 auto-connection is shown.)
2. No pulse voltage appears on the cathode of the efficiency diode.

Disadvantages:

1. D.C. flows in coil with auto-connection unless capacitor is used.
2. Tertiary winding has negative pulse in it and so very large voltage across transformer.
3. Poor coupling between primary and tertiary is likely and rings will be present.



④

Fig. 23.

4. The efficiency diode has the boosted potential between heater and cathode.

(4) *Efficiency Diode in Secondary; Boost added to H.T. Positive line.* (*Fig.* 23.4)

The efficiency diode in series with a capacitor is connected across the deflector coils. The boost capacitor is charged to approximately the potential across the deflector coils during the scanning stroke. This boost potential is added to the positive H.T. line by the method of connection shown in Fig. 23.4. It should be noted that boost potential depends on the losses in the circuit and not on the actual potential which appears across particular coils. If the obtainable boost potential and the potential across the coils during the scanning stroke are not compatible, adjustment is made by tapping the coils into the secondary winding at a suitable point.

Advantages:

1. No pulse voltage appears on the diode cathode.
2. Extra H.T. potential is available for other parts of the receiver.
3. The pulse on the secondary can be utilized by the E.H.T. circuit in addition to the primary pulse.

Disadvantages:

1. Auto-transformer connection cannot be used.
2. D.C. flows in the secondary winding.
3. The D.C. heater-to-cathode potential of the efficiency diode is the boosted potential.

It should be noted that the lower the power losses in the various components, particularly the deflector coils, the more important is leakage inductance in the transformer. These leakage inductances ring with the stray capacitance and produce velocity modulation of the scan. Auto-transformer connection permits better coupling and is more advantageous the higher the impedance of the deflector coils.

Appendix II: Scanning Coils for 70 deg. Deflection Angle

Yoke

Two Ferroxcube castellated rings (Type FX1137/1).

Windings

Short coils (line) two coils connected in series. Each coil 253 turns of ·0116 in. diameter (31 S.W.G.) copper wire covered with enamel and single silk.

Long coils (frame) two coils connected in series. Each coil 830 turns of ·0092 in. diameter (34 S.W.G.) copper wire covered with enamel.

Assembly ·

All coils are wound on a former $2\frac{1}{2}$ in diameter with cheeks $\frac{1}{4}$ in. apart. They are then formed with parallel sides and the parallel sides taped. The two yoke rings are held together and the four coils inserted. A S.R.B.P. spacer ring ($\frac{1}{2}$-in. width) is inserted to space the yoke rings and carry the soldering tags. The ends of the coils are then bent up and tied in position. The coils are connected to the appropriate tags and the damping resistors 3·3 kΩ connected across each frame coil. Finally, each end of the assembly is dipped in a suitable protective coating material. Fig. 24 shows completed assembly.

Electrical Specification:
Line coils (short)

Inductance 32·25 mH
Resistance 29 Ω
Deflection sensitivity D 66·7 cms/A, i.e., 545 mA required for 38 cms deflection on Mullard MW41-1 with E.H.T. 14 kV.

$$\therefore \frac{D}{\sqrt{L}} = 12{\cdot}3$$

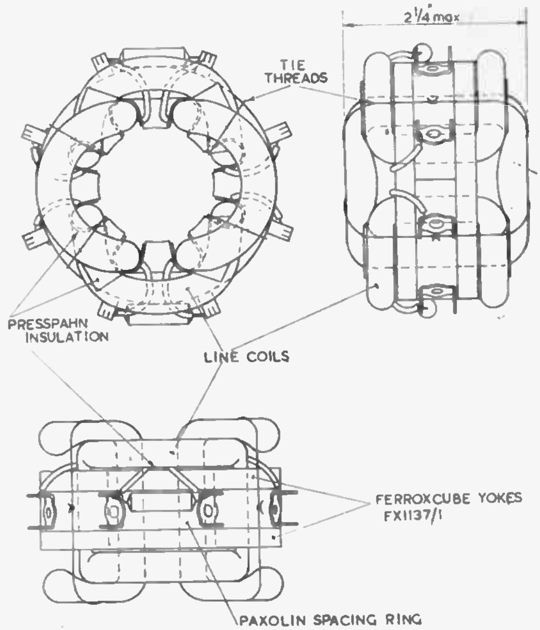


Fig. 24.—Experimental scanning coil using a castellated "Ferroxcube" yoke.

Frame Coils (Long)

The following data includes the effect of the 3·3-KΩ resistors connected across each coil.

Resistance R	148 Ω
Inductance L	287 mH
Deflection Sensitivity D	219 cms/amp.

i.e. 128 mA for 28 cm deflection on Mullard MW41-1 with E.H.T. 14 kV.

$$\therefore \frac{D}{\sqrt{R}} = 15{\cdot}3$$

Appendix III: Optimum Value of Transformer Secondary Inductance

Fig. 25 shows the equivalent circuit.

The voltamps applied to AB are $V_1\,(I_1 + I_2)$.

Now $V_1 = V_2 \dfrac{\sigma L_1 + L_2}{L_2} = V_2(1 + \sigma/p)$

where $p = L_2/L_1$

and $I_1/I_2 = (L_2 + \sigma L_1)/L_1 = p + \sigma$

hence $(I_1 + I_2)/I_2 = 1 + p + \sigma$

Thus

$$\frac{\text{Voltamps applied to AB}}{\text{Voltamps in } L_2} = V_1(I_1 + I_2)/V_2 I_2$$
$$= (1 + \sigma/p)(1 + \sigma + p) \quad \text{..........(1)}$$

For this ratio to be a minimum we have

$$\frac{d}{dp}\left[(1 + \sigma/p)(1 + \sigma + p)\right] = 0$$

from which one obtains

$$p^2 = \sigma^2 + \sigma \quad \text{......................(2)}$$

Inserting

$$\sigma = (1 - K^2)/K^2 \text{ whence } 1 + \sigma = 1/K^2$$

and writing

$$p^2 = \sigma(1 + \sigma)$$

we get $p^2 = (1 - K^2)/K^4$

or $L_2/L_1 = p = (1 - K^2)^{\frac{1}{2}}/K^2$ ....(3)

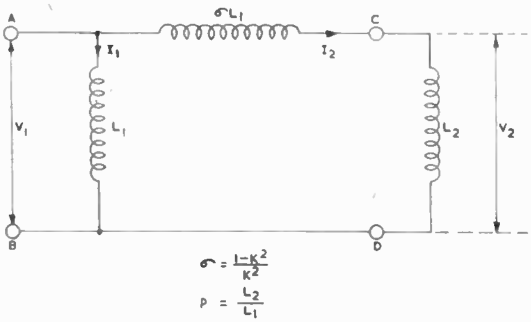


Fig. 25.—Circuit for analysis of energy relations.

Hence for max. voltamps transfer

$$L_1 = [K^2/(1 - K^2)^{\frac{1}{2}}]\,L_2 \quad \text{............(3a)}$$

Inserting (3) in (1) gives

$$\frac{\text{Voltamps in AB}}{\text{Voltamps in } L_2} = [(1 + \sigma) + p]\,[1 + \sigma/p]$$
$$= [1/K^2 + (1 - K^2)^{\frac{1}{2}}/K^2]\,\{1 + [(1 - K^2)/K^2]\,[K^2/(1 - K^2)^{\frac{1}{2}}]\}$$
$$= [1 + (1 - K^2)^{\frac{1}{2}}]^2/K^2 \quad \text{....(4)}$$
$$\text{or } = [2 - K^2 + 2(1 - K^2)^{\frac{1}{2}}]/K^2 \quad \text{................(4a)}$$

Impedance across AB =

$$V_2[1 + (1 - K^2)^{\frac{1}{2}}]/I_2\{[1 + (1 - K^2)^{\frac{1}{2}}]/K^2\}$$
$$= K^2\,V_2/I_2$$

i.e. the impedance is altered only by the factor K^2, the current increasing by the factor $[1 + (1 - K^2)^{\frac{1}{2}}]/K^2$ while the voltage increases by the factor $1 + (1 - K^2)^{\frac{1}{2}}$.

If L_1 is reduced below the optimum value to reduce the energy in the leakage inductance the voltamps input is increased. Let us calculate the amount of this increase as a function of the ratio α of L_1 to L optimum. Thus, if $L_{opt} = [K^2/(1 - K^2)^{\frac{1}{2}}]L_2$ ....(3c)

and $L_1 = \alpha\,L_{opt}$ ....(5)

we have $V_1(I_1 + I_2)/V_2 I_2 = (1 + \sigma\,L_1/L_2)(1 + L_2/L_1 + \sigma)$ from (1).

Inserting $L_1/L_2 = \alpha\,K^2/(1 - K^2)^{\frac{1}{2}}$ from (3) and (5) we get

$$V_1(I_1 + I_2)/V_2 I_2 =$$
$$\{1 + [(1 - K^2)/K^2][\alpha K^2/(1 - K^2)^{\frac{1}{2}}]\}\,\{1 + (1 - K^2)^{\frac{1}{2}}/\alpha K^2 + (1 - K^2)/K^2\}$$

which simplifies to

$$[2\alpha + (1 - K^2)^{\frac{1}{2}}(1 + \alpha^2) - K^2\alpha]/\alpha K^2$$

$$\text{or } \frac{\text{voltamp input}}{\text{voltamps in } L_2} = \frac{2 - K^2}{K^2} + \left(\frac{1 + \alpha^2}{\alpha}\right)\frac{\sqrt{1 - K^2}}{K^2} \quad \text{......................(6)}$$

When $\alpha = 1$ this becomes identical with (4*a*) which it should. A plot of $\frac{\text{Total V.A.}}{\text{Load V.A.}}$ against K and α is given in Fig. 8.

The energy stored in the series reactance is given by

$$I_2(V_1 - V_2) = V_2\sigma I_2/p = (\sigma/p)(V_2 I_2)$$

writing

$$p = L_2/L_1;\ \sigma = (1 - K^2)/K^2;\ \alpha = L_1/L_{opt};\ L_{opt} = K^2L^2/(1 - K^2)^{\frac{1}{2}}$$

we have

$$\frac{\text{Series V.A.}}{V_2 I_2} = \sigma/p = [(1 - K^2)/K^2][K^2\alpha/(1 - K^2)^{\frac{1}{2}}]$$
$$= \alpha/(1 - K^2)^{\frac{1}{2}} \quad \text{....................(7)}$$

which is also plotted in Fig. 8.

One can also derive

$$\frac{\text{Parallel V.A.}}{V_2 I_2} = (p + \sigma)^2/p$$
$$= (1 - K^2)^{\frac{1}{2}}[1 + \alpha(1 - K^2)^{\frac{1}{2}}]/\alpha K^2 \quad \text{................(8)}$$

OBSTACLE-TYPE ARTIFICIAL DIELECTRICS FOR MICROWAVES*

by

Charles Süsskind, Ph.D. (*Associate Member*)†

A Paper presented at the Third Session of the 1951 Radio Convention on July 24th at University College, Southampton.

SUMMARY

Artificial dielectrics consisting of arrays of small metallic elements have found widespread employment recently, notably in the construction of microwave lens antennæ. Depending on the shape of the obstacles, these media may be conveniently analyzed (1) by an analogy with classical molecular theory, (2) by a consideration of scattering phenomena, or (3) by an analogy with transmission-line theory. A great deal of experimental information is also becoming available.

A survey of the analytical and practical techniques now used in the design of artificial dielectrics is presented. Special emphasis is given to a consideration of the similarities and dissimilarities between the behaviour of such media and that of actual dielectrics. The case for artificial dielectrics is discussed in terms of their various advantages over actual dielectrics, such as light weight, good matching properties, and the feasibility of spatial variation of refractive index for Lüneburg-lens and other applications.

1. Introduction

Artificial dielectrics have recently come into extensive use, particularly in connection with the construction of lens antennæ and other configurations intended for the microwave frequency range. The principal effect of *actual* dielectrics on electromagnetic radiation is to alter its velocity of propagation. An *artificial* dielectric is a device which can simulate this effect. One of the first types employed was the waveguide or "metal-plate" artificial dielectric. This medium consists essentially of an array of waveguides or merely parallel metal plates, within which the phase velocity is greater than in free space; the effective refractive index is thus smaller than unity, and appropriate configurations can be used for phase correction of microwave horns or for lenses. The medium was developed independently in Britain,[33] America,[19] and Germany.[35] A lens made from it is cheaper and lighter than an equivalent lens made of glass or polystyrene, and may be designed to withstand more abuse. It has certain disadvantages, the chief among which are that lens action is restricted to a single polarization and to a narrow frequency band.

The employment of *obstacle-type* artificial dielectrics, first proposed by Kock,[20] does away with both of these difficulties. An elementary form comprises a cubic lattice of metal spheres having a radius and separation which are small compared with the wavelength. The array thus seeks to reproduce the molecular structure of an actual dielectric on a macroscopic scale. The submicroscopic particles which (in an actual dielectric) may be thought of as "delaying," say, light waves, are replaced by metallic obstacles of millimetric dimensions, which act similarly to delay waves of centimetric lengths. As in the case of actual dielectrics, the delaying action takes place only if the wavelength is large compared with the spacing between the particles. This fact introduces a frequency limitation after all, though not nearly as severe a one as that of the metal-plate medium. The refractive index begins to rise as the frequency is increased, until the medium becomes opaque to radiation. At still higher frequencies, the refractive index drops abruptly, in analogy with the anomalous dispersion effect present in actual dielectrics, and then rises again and approaches unity.

As a result of these analogies, the medium can be analyzed approximately on the basis of classical molecular theory, as shown in Section 2. The basic theory assumes that each particle is polarized by an incident field, and that the total polarization, from which the properties of the

* Manuscript received July, 1951.

† Stanford University, Stanford, California. (This paper, which was prepared during the author's tenure at Yale University, New Haven, Connecticut, and is based in part on a Ph.D. dissertation at Yale University, is the outcome of an investigation sponsored by the U.S. Air Force Cambridge Research Centre, Cambridge, Massachusetts, under the contract described in Reference 9.)

U.D.C. No. 621.396.677.

medium may be deduced, is the sum of the effects of the individual particles. A more exact approximation takes into consideration the effect of the surrounding particles in *altering* the field which excites an individual particle, and leads to the Clausius-Mossotti equation of molecular theory. A still more sophisticated approach, which proceeds from scattering considerations, is described in Section 4; it is gratifying to note that the results of this analysis reduce to the Clausius-Mossotti formula as a limiting case.

In his original analysis, Kock pointed out[20] that obstacles which extend in the direction of propagation (i.e., spheres and all other "thick" obstacles) will disturb the magnetic field in such a way as to produce a diamagnetic effect; the bulk relative permeability of a medium made up of such obstacles will be less than unity. The bulk refractive index, given by the square root of the product of the dielectric constant and the relative permeability, will thus be reduced below the value which would be attained if this effect were not present. This difficulty can be obviated by replacing the metal obstacles by dielectric obstacles, or by using metal obstacles which are *thin* in the direction of propagation, such as discs or strips. If strips are used, waves of a single polarization only will be affected (E normal to the strip length). Despite this limitation, the strip medium has found widespread employment (principally because of ease in construction), and a more exact analysis had to be developed. Such an analysis, based on an analogy with transmission-line theory, is described in Section 3. With certain reservations, this method can be extended to the analysis of media containing discs as well. Again, the results appear to reduce to the molecular-theory formulae in the limit.

One of the advantages of obstacle-type artificial dielectrics is that they may be conveniently constructed to provide a good match to free space. In optics, matching between two dielectrics is obtained by inserting a third medium which has a characteristic impedance equal to the geometric mean of the two characteristic impedances to be matched, and a thickness equal to a quarter wavelength (within the matching medium). The "impedance" terminology follows readily from the transmission-line analogy: a corresponding device (a quarter-wavelength transformer) can be reproduced quite easily with artificial dielectrics simply by reducing the size of the obstacles in the outermost layers. This procedure, first suggested by Cohn for the strip medium,[6] makes it possible to obtain a match to free space over a wide frequency band.

Still another advantage of obstacle-type artificial dielectrics is that they can be readily made to have a refractive index which varies spatially. (This characteristic cannot be easily achieved with actual dielectrics, although some attempts have been made to use expanded plastics compressed to varying extents.) The refractive index can be made to vary according to a prescribed law by using obstacles whose spacing or separation (or both) are varied continuously. In this way, a lens employing the Lüneburg principle[25] can be constructed; such a lens has many applications in radar scanning.

Artificial dielectrics comprising discrete obstacles are usually constructed with the aid of an expanded plastic material as the embedding medium. Styrofoam (polystyrene foam) and Plasticell (expanded polyvinyl chloride) are among the materials which have been used for this purpose. The expanded plastics are characterized by an extraordinarily low density and by dielectric properties closely approximating those of air. The fact that a fairly rigid expanded plastic is not generally available in some countries (e.g., Britain) is one of the reasons why obstacle-type artificial dielectrics have not received as much attention as the older waveguide or metal-plate types and their variants, such as the path-length medium[21] and the "rodded" medium.[3]

The application to microwave lenses by no means exhausts the uses to which artificial dielectrics of the obstacle type can be put. They may be also used in many other ways, such as attenuators and non-reflecting walls. Moreover, they can be utilized in developing equivalent structures which simulate the characteristics of various waveguide and antenna configurations (e.g., polyrod antennæ), and new uses are continually being discovered.

2. Molecular-Theory Approach

2.1. *Basic Theory*

The most simple calculation by which the bulk dielectric constant, relative permeability, and refractive index of an artificial dielectric containing metal obstacles may be obtained is based on the assumption that each obstacle acquires electric and magnetic dipole moments under the action of an incident field. The analysis then proceeds in a manner similar to that utilized in computing the properties of an actual dielectric, with the obstacles taking the place of nonpolar, diamagnetic, atoms or molecules. (This representation first led to the concept of artificial dielectrics of the obstacle type.) This approach comes to mind most readily because it affords a direct analogy with actual dielectrics, and because classical molecular theory is largely based on mechanical models which bear a close resemblance to an artificial dielectric containing, say, spherical obstacles. Thus, a gas molecule is pictured as a conducting sphere of radius a, which becomes polarized under the influence of an external electric field of magnitude E, i.e., acquires a dipole moment

$$m_e = \alpha_e E \quad \text{(1)}$$

where α_e is the electric polarizability. If there are N molecules per unit volume, the total dipole moment per unit volume, called *polarization*, is

$$P = N\alpha_e E \quad \text{(2)}$$

It is customary to define a displacement vector of magnitude

$$D = \varepsilon E = \varepsilon_0 E + P \quad \text{(3)}$$

where ε is called the *permittivity*; its ratio to the permittivity of free space,

$$\kappa_e = \frac{\varepsilon}{\varepsilon_0} \quad \text{(4)}$$

is called the *dielectric constant*. In M.K.S. units, $\varepsilon_0 = 8{\cdot}85 \times 10^{-12}$ farad/metre.

The ratio of P to $\varepsilon_0 E$ is the *electric susceptibility* and is denoted by χ_e; thus

$$\chi_e = \frac{P}{\varepsilon_0 E} = \frac{\alpha_e}{\varepsilon_0} N \quad \text{(5)}$$

so that finally

$$\kappa_e = 1 + \frac{P}{\varepsilon_0 E} = 1 + \frac{\alpha_e}{\varepsilon_0} N \quad \text{(6)}$$

The expression for relative permeability is derived similarly; it is given by

$$\kappa_m = \frac{\mu}{\mu_0} = 1 + \frac{\alpha_m}{\mu_0} N \quad \text{(7)}$$

where $\mu_0 = 4\pi \times 10^{-7}$ henry/metre is the permeability of free space in M.K.S. units.

The diamagnetic properties alone are considered here. Even if ferromagnetic particles are used, it is found that they behave like diamagnetic particles at microwave frequencies. The magnetic polarizability will thus be *negative* for all conducting obstacles except for infinitely thin discs or strips, arranged with their flat sides parallel to the H-field of a wave propagated in a direction normal to the flat sides; in this case, $\alpha_m = 0$ and the relative permeability is unity.

The electric and magnetic polarizabilities of various obstacle shapes, obtained by several investigators, are tabulated below. The polarizabilities may be obtained from classical static theory, or by the consideration of the torque on an obstacle in a uniform field. Moreover, the number of available expressions has been significantly extended by Cohn, who pointed out[3] that the simple correspondence between the electric and magnetic polarizability of a single conducting disc on one hand, and the magnetic and electric polarizability of an aperture (of the same shape as the disc) in a conducting wall on the other hand, make it possible to write down many of the formulæ directly (Babinet's principle). The reference listed for some of these obstacles[29] is actually a tabulation of aperture polarizabilities. The sources listed for the other obstacles do not refer in every case to original computations, but rather to the first mention of the obstacle under consideration in connection with artificial-dielectric theory. Unit-length values are given for the cylinders and strips.

The refractive index of the medium is given by

$$n = \sqrt{\kappa_e \kappa_m} \quad \text{(8)}$$

and since κ_m is less than unity for artificial dielectrics containing thick conducting obstacles, the range of n is quite limited. Furthermore, thick obstacles cannot be packed as closely as thin obstacles, so that the values of n attainable with thick obstacles are limited by physical considerations. From the point of view of obtaining high values of refractive index, thin obstacles are therefore to be preferred. On the other hand, the use of thin obstacles introduces an obvious anisotropy. This anisotropy is of no consequence if only plane waves propagating along the medium axis need be considered. However, for applications demanding a higher degree of isotropy, thick obstacles must be used in spite of the consequent restriction of the refractive index to low values. Even then, some anisotropy remains because of the nature of the array, especially if the cross-section of the incident beam is small.

TABLE 1

Electric polarizabilities of various obstacles

Obstacle	α_e/ε_0	Source
Sphere (rad. a)	$4\pi a^3$	20
Cylinder (rad. a, $E \perp$ axis)	$2\pi a^2$	32
Thin obstacles ($E \parallel$ face):		
Circular disc (rad. a)	$16a^3/3$	20
Ellipse* ($E \parallel b$)	$\frac{4\pi}{3}\,\frac{ab^2e^2}{[E(e)-(1-e^2)F(e)]}$	29
Ellipse ($E \parallel a$)	$\frac{4\pi}{3}\,\frac{ab^2e^2}{(1-e^2)[F(e)-E(e)]}$	29
Long narrow ellipse ($a \gg b$, $E \parallel a$)	$\frac{4\pi}{3}\,\frac{a^3}{\log_e(4a/b)-1}$	29
Long narrow ellipse ($a \gg b$, $E \parallel b$)	$4\pi ab^2/3$	29
Strip (width w, $E \parallel$ length)	$\pi w^2/4$	20
Square (side d)	$1{\cdot}032\, d^3$	8

* Ellipse has semi-axes a (major) and b (minor) and eccentricity $e = \sqrt{1-(b/a)^2}$. The functions $F(e)$ and $E(e)$ are the complete elliptic integrals of the first and second kind, respectively, and are tabulated by Jahnke and Emde.[17] They are given by

$$F(e) = \int_0^{\pi/2} \frac{du}{\sqrt{1-e^2\sin^2 u}} \qquad E(e) = \int_0^{\pi/2} \sqrt{1-e^2\sin^2 u}\, du$$

TABLE 2

Magnetic polarizabilities of various obstacles

Obstacle	α_m/μ_0	Source
Sphere (rad. a)	$-2\pi a^3$	20
Cylinder (rad. a, $H \parallel$ axis)	$-\pi a^2$	32
Cylinder (rad. a, $H \perp$ axis)	$-2\pi a^2$	9
Disc of moderate thickness (volume V, $H \parallel$ face)	$-V$	8
Thin obstacles ($H \perp$ face):		
Circular disc (rad. a)	$-8a^3/3$	13
Ellipse (*cf.* footnote, Table 1)	$-\frac{4\pi}{3}\,\frac{ab^2}{E(e)}$	29
Long narrow ellipse ($a \gg b$)	$-4\pi ab^2/3$	29
Strip (width w)	$-\pi w^2/4$	29
Square (side d)	$-0{\cdot}455\, d^3$	9

2.2. *The Clausius-Mossotti Formula*

The simple expressions derived in the preceding sub-section will hold exactly only if the spacing between the obstacles is very large compared with their dimensions. Returning to the analogy with molecular theory, it may be said that the relationships developed for gases will not give correct results for the more dense media. For instance, in calculating the polarization from (2) for a liquid or solid, it must be realized that the polarization is no longer merely that caused by the external field acting on a virtually isolated molecule, but also that due to the attractions and repulsions by neighbouring molecules (themselves polarized by the external field).

The effect of the neighbouring particles on the value of the exciting field is found by a computation which may be found in many standard texts on electromagnetic theory. The result for the electric case is that the exciting field is given by the external field plus $P/3\varepsilon_o$, so that instead of (2) there is

$$P = N\alpha_e\left(E + \frac{P}{3\varepsilon_o}\right)$$

or

$$P = \frac{N\alpha_e E}{1 - \dfrac{N\alpha_e}{3\varepsilon_o}} \quad \ldots\ldots\ldots\ldots\ldots(9)$$

The susceptibility χ_e is the ratio of P to ε_o times the total field:—

$$\chi_e = \frac{P}{\varepsilon_o E} = \frac{\dfrac{N\alpha_e}{\varepsilon_o}}{1 - \dfrac{N\alpha_e}{3\varepsilon_o}} \quad \ldots\ldots\ldots\ldots\ldots(10)$$

and the dielectric constant is

$$\kappa_e = 1 + \chi_e = 1 + \frac{\dfrac{N\alpha_e}{\varepsilon_o}}{1 - \dfrac{N\alpha_e}{3\varepsilon_o}} = \frac{1 + \dfrac{2}{3}\dfrac{N\alpha_e}{\varepsilon_o}}{1 - \dfrac{1}{3}\dfrac{N\alpha_e}{\varepsilon_o}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

This "corrected" value of κ_e takes into account the effect of the surrounding polarized particles in changing the field. The expression for κ_e is often written

$$\frac{\kappa_e - 1}{\kappa_e + 2} = \frac{N\alpha_e}{3\varepsilon_o} \quad \ldots\ldots\ldots\ldots\ldots\ldots(12)$$

and is known as the *Clausius-Mossotti* formula.

An analogous formula is obtained for relative permeability:

$$\frac{\kappa_m - 1}{\kappa_m + 2} = \frac{N\alpha_m}{3\mu_o} \quad \ldots\ldots\ldots\ldots\ldots\ldots(13)$$

These expressions can also be written in terms of the fractional volume f occupied by the obstacles: for instance, for *spherical* conducting particles (where $\alpha_e = 4\pi a^3\varepsilon_o$ and $\alpha_m = -2\pi a^3\mu_o$), which occupy a fractional volume

$$f = \frac{4}{3}\pi a^3 N$$

the expression (11) for dielectric constant becomes

$$\kappa_e = \frac{1 + 2f}{1 - f} \quad \ldots\ldots\ldots\ldots\ldots\ldots(14)$$

and the relative permeability is

$$\kappa_m = \frac{1 - f}{1 + f/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(15)$$

so that

$$n = \sqrt{\frac{1 + 2f}{1 + f/2}} \quad \ldots\ldots\ldots\ldots\ldots(16)$$

If the obstacles are contained in an embedding medium of constants κ_e' and κ_m', the right-hand sides of (14) and (15) must be multiplied by κ_e' and κ_m', respectively; and the right-hand side of (16) must be multiplied by $\sqrt{\kappa_e'\kappa_m'}$.

The above analysis can be readily extended to *dielectric* spheres embedded in another dielectric. In this case, $\kappa_m = 1$; and if the spheres themselves are made of a dielectric material having a dielectric constant κ_e'', it can be shown that (14) becomes

$$\kappa_e = \kappa_e' \frac{1 + 2f\,\dfrac{\kappa_e'' - \kappa_e'}{\kappa_e'' + 2\kappa_e'}}{1 - f\,\dfrac{\kappa_e'' - \kappa_e'}{\kappa_e'' + 2\kappa_e'}} \quad \ldots(17)$$

Thus the substitution of dielectric obstacles for conducting ones will have the effect of reducing the bulk dielectric constant, since the factor $(\kappa_e'' - \kappa_e')/(\kappa_e'' + 2\kappa_e')$ is smaller than unity for all $\kappa_e'' > \kappa_e'$. On the other hand, κ_m has been increased to unity as a result of zero eddy currents, and the value of n will be actually increased, provided κ_e'' exceeds a certain value (about $3\kappa_e'$).

The behaviour of n as a function of s/a, the spacing-to-radius ratio, is shown in Fig. 1 for several values of κ_e'' (based on air as the ambient medium), for a cubic array of spheres. The plot for conducting obstacles (dashed line) is shown for comparison.

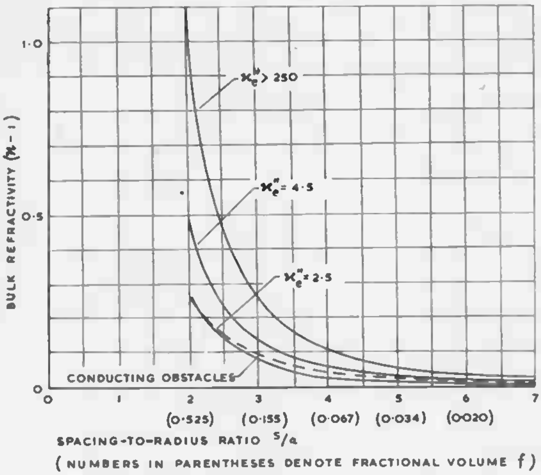


Fig. 1.—Bulk refractivity vs. *spacing-to-radius ratio for cubic array of spheres.*

The refractive index for dielectric obstacles approaches a limiting value

$$\lim_{\kappa_e'' \to \infty} n = \sqrt{\frac{1+2f}{1-f}}$$

Actually, n assumes about 95 per cent. of this limiting value when κ_e'' is only 25, and is virtually indistinguishable from the limiting value at $\kappa_e'' = 250$. Thus very little would be gained by using spheres having a very high dielectric constant (e.g., the new titanates) in place of dielectrics having a moderately high value of κ_e''—to say nothing of the probable increase in attenuation inherent in the use of the high-κ_e'' dielectrics.

3. Transmission-Line Approach

3.1. *The Strip Medium*

Metal strips, extending at right-angles to the direction of propagation in such a way that the E-vector is parallel to the strip faces and perpendicular to the strip lengths, constitute a very popular type of artificial dielectric. (With reference to Fig. 2 (*a*), the incident wave would have a vertical E field and propagate to the right.) When the successive strip layers are stacked closely, it is found that the formulæ of Section 2 no longer apply.[18] An alternative theory, based on an analogy with a loaded transmission line, has been proposed almost simultaneously in America[6] and Britain.[2] If the E-vector of a normally incident wave is oriented as described above, thin conducting sheets can be inserted, as shown in Fig. 2 (*a*) without affecting the field. The resulting structure, an elementary waveguide shunted down by successive capacitive irises, is shown in Fig. 2 (*b*). The equivalent circuit is shown in Fig. 2 (*c*); the normalized susceptance B is given approximately by

$$B = \frac{4s}{\lambda} \log_e \operatorname{cosec} \frac{\pi(s-w)}{2s} \quad \ldots\ldots\ldots (18)$$

where s is the spacing (on centres) between strips in the same layer, and w is the strip width. The refractive index of the medium is

$$n = \sqrt{1 + \frac{B\lambda}{2\pi l}} \quad \ldots\ldots\ldots (19)$$

where l is the spacing between successive strip layers.

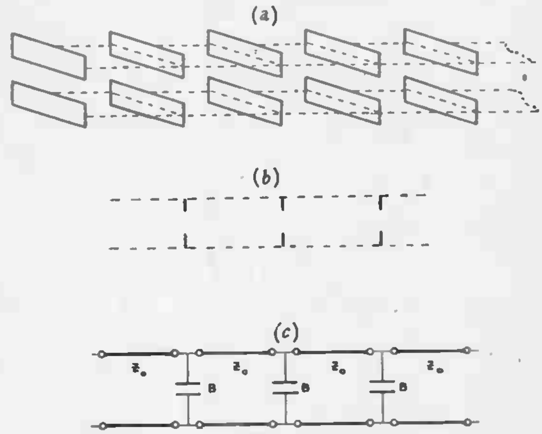


Fig. 2.—Transmission-line equivalent of metal-strip dielectric.

Equations (18) and (19) are valid only if the irises can be considered to be isolated (if $l/s > 0{\cdot}75$, say), and if the strip separation is small compared with the wavelength ($s/\lambda < 0{\cdot}2$). For larger s, a more exact formula for B must be used. If the successive strip layers are very close together (*i.e.*, if $l/s < 0{\cdot}75$), the value of B is decreased, and "bridging" capacitances appear between adjacent sections. These capacitances are shown by dotted lines in the equivalent circuit of Fig. 3, in which interaction between nearest neighbours only is assumed. The analysis of this circuit is quite difficult, but the results are available in the form of graphs and tables.[6] The series inductance of Fig. 3 has been introduced to take into account the finite thickness of the obstacles.[9]

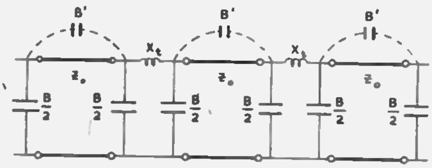


Fig. 3.—Exact equivalent circuit for metal-strip dielectric.

3.2. *Discrete Discs*

The analysis of the preceding sub-section may be applied to disc-shaped particles as well. However, the susceptance involved is no longer that of a simple iris, and cannot be computed so readily. An ingenious method of determining this susceptance *experimentally* with the aid of an electrolytic tank[7] has been developed by Cohn.[8,9] The susceptance of a waveguide "cell" is measured by this analogue with and without the obstacle; the susceptance of disc obstacles of arbitrary shape and moderate thickness (and hence the bulk refractive index of the medium) may be thus determined statically. The behaviour of the refractive index at *high* frequencies can likewise be determined experimentally. In one method, a microwave interferometer is used, in which the signal radiated from a horn is intercepted by a receiver horn and mixed with a second signal, obtained directly from the microwave source via a constant-length path. A "null" is obtained by adjusting the position of the receiver horn; the artificial-dielectric sample is next inserted between the horns, and from the change of receiver-horn position (measured by a micrometer) necessary to re-establish the null, the phase shift experienced by the wave due to the presence of the sample is determined. The phase shift and the known sample thickness yield the bulk refractive index. This technique has also been perfected by Cohn,[10] who has developed a procedure which takes into account the various sources of error arising in the measurement (notably the effect of the surface discontinuity, discussed in Section 5 below).

The results of the static (electrolytic-tank) measurements[7] for circular and square discs have been replotted in Fig. 4 in terms of the susceptibility $\chi_e = N\alpha_e/\varepsilon_o$, where α_e/ε_o is given for the various obstacles by the appropriate expression from Table 1. This plot affords a direct comparison with the values of refractive index obtained from the Clausius-Mossotti formula, which for normal incidence now takes the form (in free space)

$$n = \sqrt{\kappa_e} = \sqrt{\frac{1 + 2\chi_e/3}{1 - \chi_e/3}} \quad \ldots\ldots\ldots\ (20)$$

since in this case, $\kappa_m = 1$.

The dashed curves of Fig. 4 are plots of Equation (20), and are of course identical in all cases. The experimental values, which are plotted for various values of the l/s ratio, are shown to approach the Clausius-Mossotti curve as this ratio increases towards unity. It should be

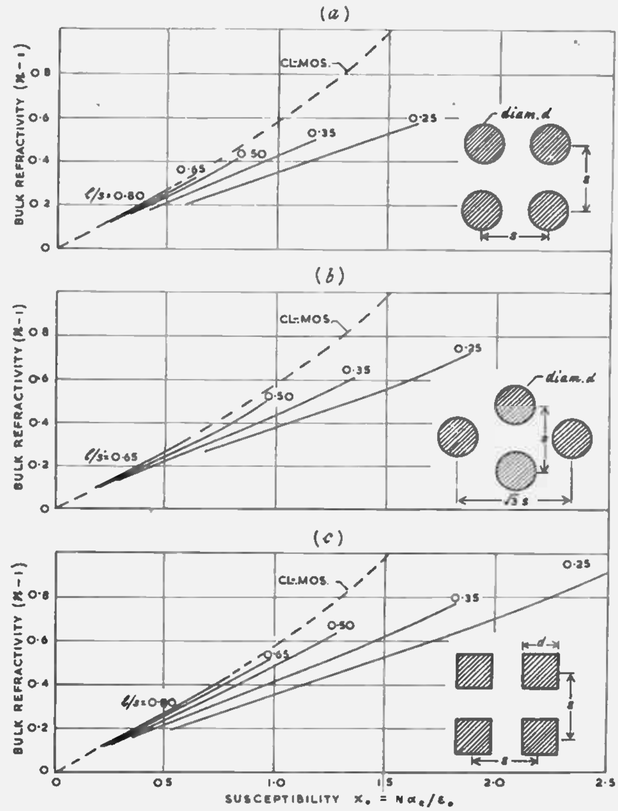


Fig. 4.—Measured bulk refractivity vs. *calculated susceptibility for discs.*

pointed out that this gratifying result would perhaps not be obtained if the "fractional area" taken up by the discs in any one layer approached the maximum value of unity, or if the l/s ratio became large (say, larger than unity). These two cases are, however, of comparatively little practical interest.

A drawback of the transmission-line approach for discs is the fact that the susceptance of the various obstacles must be obtained experimentally. A possible (entirely theoretical) approach is discussed at the end of the next section. Finally, it should be recalled that all thin-obstacle media are of course anisotropic. The relative permeability will fall below unity for *oblique* incidence, and reach a minimum when the obstacle face lies perpendicular to the H-vector. This case has been considered (for circular discs) at some length by Estrin.[13, 14]

4. Scattering-Theory Approach

4.1. *Spherical Obstacles*

A rigorous method of analyzing the bulk properties of a cubic array of spheres was published by Lewin[22] even before the application of such arrays to microwave lenses had been proposed. Lewin's analysis begins with a computation of the field scattered by a sphere. (This computation is based on that made in 1908 by Mie[27]; the subsequent work on this topic has been reviewed most adequately by a Dutch astronomer, H. C. van de Hulst,[16] who is now preparing a book on the subject.) The computation is amended to take into account the effect of the surrounding spheres, and the resulting expression is summed over a semi-infinite array. Two expressions are obtained, for the electric and for the magnetic case, and from a consideration of the reflected and transmitted waves, κ_e and κ_m are obtained by analogy with a semi-infinite slab of actual dielectric. The scattered-field equations assume a comparatively simple form if the sphere dimensions are small compared with the wavelength *in the ambient medium*. For this case, the results resemble the equations of Section 2 in form: for instance, the equation for bulk dielectric constant becomes

$$\kappa_e = \kappa_e' \frac{1 + 2f\dfrac{\kappa_{ep} - \kappa_e'}{\kappa_{ep} + 2\kappa_e'}}{1 - f\dfrac{\kappa_{ep} - \kappa_e'}{\kappa_{ep} + 2\kappa_e'}} \quad \ldots\ldots\ldots(21)$$

where κ_{ep} is related to the dielectric constant κ_e'' of the sphere material as follows:

$$\frac{\kappa_{ep}}{\kappa_e''} = \frac{2(\sin\theta - \theta\cos\theta)}{(\theta^2 - 1)\sin\theta + \theta\cos\theta} \quad \ldots\ldots\ldots(22)$$

where

$$\theta = k_2 a = \frac{2\pi}{\lambda} a\sqrt{\kappa_e''\,\kappa_m''} \quad \ldots\ldots\ldots(23)$$

and similarly for κ_{mp}. The rather complicated expression (22), which specifies the dependence of κ_e on frequency, reduces to unity if the sphere dimensions are small compared with the wavelength *in the material of the spheres*: in this case θ is small, and (22) expands according to

$$1 + \frac{\theta^2}{10} + \frac{9\theta^4}{700}$$

so that (21) reduces to the Clausius-Mossotti equation (17). In the case of conducting obstacles, $\kappa_{ep} \gg \kappa_e'$ and (at microwave frequencies) $\kappa_{mp} \rightarrow 0$; the reader can readily verify that the expression for the bulk refraction index of an array of conducting obstacles then becomes identical with Equation (16) of the molecular-analogy approach of Section 2.

(A mathematically similar problem has been considered by Page in connection with the determination of the magnetic properties of permeable spheres at metric wavelengths.[31])

4.2. *Disc Obstacles*

The inadequacies of both the molecular-theory and transmission-line analogies in the analysis of disc-obstacle arrays have been mentioned above: the former agrees with experiment only over a limited range, and the latter introduces a measured quantity into the computation. It is possible that a satisfactory theory might be derived from a combination of some procedure, such as that described in the preceding subsection, with the mutual-coupling concepts of transmission-line theory. The mathematics would presumably be that concerned with oblate spheroids (of which the circular disc is a special case). The available literature on this subject is quite voluminous, ranging from the classic work of Möglich[30] to some excellent papers published recently in America,[24] France,[34] Germany,[26] and Holland.[1] The case of discs of other than circular shape might then follow on the basis of a consideration of the fractional area occupied by the discs in any one layer, with due regard for the additional anisotropy introduced by shapes such as rectangles or ellipses. The results presented in Section 3 would provide an indirect check on any equation obtained by such a method—i.e., it would have to reduce to the Clausius-Mossotti formula (17) in the limit.

5. The Interface Problem

The equations presented above are based on the assumption that the artificial-dielectric medium is of infinite extent. The constants of a finite slab will in general differ from those computed on the basis of an infinitely thick slab.

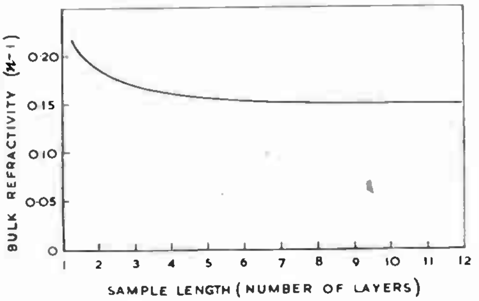


Fig. 5.—Bulk refractivity as a function of sample length

This effect also exists in actual dielectrics, mainly as a result of multiple reflexions between the two boundaries. It is much more serious in the case of artificial dielectrics, where the discontinuity at each interface contributes an additional correction as well. The effect of the interface discontinuity on the bulk refractive index is illustrated by Fig. 5. As the sample is made shorter, the measured value begins to deviate from the value calculated on the basis of the infinite-length assumption; as the sample is increased in length, the theoretical value is approached asymptotically.[11] The measured values may actually fluctuate about the smooth curve of Fig. 5 as a result of the effect of multiple reflexions within the sample; this effect is periodic with $\lambda/2$ of the incident wave.

The overall transmission and reflexion coefficients of a finite slab are given in many references;[28] the important fact to remember, as Lengyel points out,[23] is that the exponential factors occurring in these coefficients are in general themselves complex in the case of artificial dielectrics. In other words, an additional phase shift is introduced at each boundary.

One way of estimating the value of this additional phase shift is to measure the bulk properties of a single layer of obstacles and to compare them with the values predicted for a thick sample. This method has been used successfully by Cohn[10] in his measurement of artificial dielectrics comprising discs. A preliminary attempt to predict this effect theoretically has also been made (for spheres).[37] The effective refractive index n_{eff} is considered to deviate from the predicted value n as a function of the sample thickness d, according to the equation

$$n_{eff} = n + \frac{x}{d} \quad \ldots\ldots\ldots\ldots\ldots (24)$$

and x is determined by means of the scattering-theory approach of Section 4. Experimental results appear capable of being described by an equation of the form of (24) quite accurately, but more data need be obtained before the method developed for predicting the value of x can be said to be universally valid.

It can be shown that the effect of the boundary-effect correction is negligible in most practical applications (say, in the computation of lens contours). The correction will only become important when the overall dimensions are not very large compared with the spacing between the obstacles—e.g., in measurements carried out on small samples.

6. Frequency Dependence

As previously mentioned, the properties of obstacle-type artificial dielectrics resemble those of actual dielectrics in their behaviour as a function of frequency. Fig. 6 is an experimental curve (for a highly dispersive medium consisting of short metal rods) obtained by Kock,[20] which illustrates this analogy with extraordinary clarity.

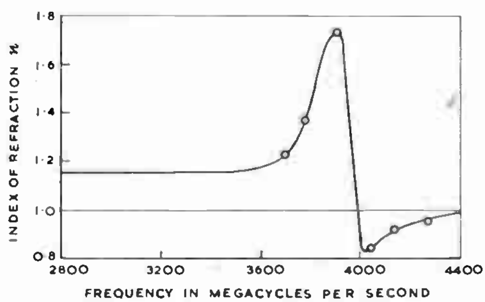


Fig. 6.—Measured refractive index of metal-rod array. (Courtesy of *The Bell System Technical Journal*).

Accordingly, the frequency dependence of a medium consisting of spheres as a function of sphere *size* could perhaps be determined on the basis of the familiar dispersion formula

$$\kappa_e - 1 = \frac{k}{\omega_o^2 - \omega^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (25)$$

Since κ_e must reduce to a static value, say κ_{edc}, when the frequency ω is very small compared with the resonant frequency ω_o, (25) may be put, after some rearrangement, in the form

$$\kappa_e - 1 = \frac{\kappa_{edc} - 1}{1 - (\omega/\omega_o)^2} \quad \ldots\ldots\ldots\ldots\ldots (26)$$

The corresponding Clausius-Mossotti equation is

$$\frac{\kappa_e - 1}{\kappa_e + 2} = \frac{1}{1 - (\omega/\omega_o)^2} \, \frac{\kappa_{edc} - 1}{\kappa_{edc} + 2} \quad \ldots\ldots (27)$$

and similarly for κ_m.

The resonant frequency for a sphere was first computed by Debye,[12] who showed that free oscillations of a perfectly conducting sphere of diameter d will take place at a number of frequencies, the lowest of which obtains when the wavelength is given by

$$kd = \sqrt{3} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28)$$

and by a slightly lower number for imperfect conductors. (The quantity $k = 2\pi/\lambda$ is the propagation constant.) For spheres, therefore, the first resonance is given by

$$\lambda_o = \frac{2\pi}{\sqrt{3}} d$$

or

$$\frac{\omega}{\omega_o} = \frac{\lambda_o}{\lambda} = \frac{3{\cdot}63\, d}{\lambda}$$

where λ is the wavelength of the incident radiation. The sphere thus reaches the first resonance point when its diameter equals between one-quarter and one-third of the incident wavelength. However, the bulk refractive index begins to deviate from the static value when the sphere diameter equals only one-tenth of the incident wavelength. On the basis of these simple considerations, the design criterion for a dispersionless medium would be that the dimensions of the obstacles should be of the order of one-tenth of the incident wavelength or smaller. Further information could probably be obtained from a detailed examination of the formulae given by Lewin,[22] who considers dielectric as well as conducting spheres.

A different sort of frequency dependence arises from the effect of particle *separation.* The case of very closely spaced particles has been considered from the scattering-theory viewpoint by Trinks,[38] who has calculated the field scattered by a sphere in the presence of a neighbouring sphere; his conclusion is that self-resonance occurs at a slightly higher value of frequency than that given by (28) for the single sphere. With the transmission-line approach (which is, after all, restricted to thin or moderately thick obstacles), a fairly complicated equivalent-circuit structure results if the separation between successive obstacle layers becomes very small (Fig. 3). The analytical difficulties of either method are very great indeed.

The obstacle medium will also become dispersive if the particle separation becomes comparable with the wavelength. A complete theory of this effect from the transmission-line viewpoint exists for the metal-strip medium.[2, 6, 15] Cohn has also extended his investigation of the strip dielectric to media comprising discrete discs[10]; the results of his measurements show that the behaviour of the disc dielectric as a function of s/λ (where s is the spacing, on centres, between discs in the same layer) is about the same for a large range of obstacle sizes, and the curve may be approximated by a relatively simple empirical formula.

It follows from the above considerations that the behaviour of obstacle-type artificial dielectrics as a function of frequency depends on both the size and the separation of the obstacles relative to the wavelength. An approximate theory for each effect has been proposed; it remains to make a study of the combination of both. The complexity of this problem may well prove to be such as to make an experimental approach the only feasible one.

7. Applications

The most extensive application of artificial dielectrics to date has been in the realm of microwave antennæ.[4, 36] This application has been adequately described in the literature, notably in publications originating at the American Telephone and Telegraph Co. (The Bell System). A typical microwave link installation[5] is the tower shown in Fig. 7. Four large horn antennæ, each corrected by an artificial-dielectric lens of the metal-strip type, are installed on the top.

Fig. 7.—Radio relay tower in microwave link system. (Courtesy of *Bell Telephone Magazine*).

The signal received by one antenna is amplified by the unattended "repeater" located in the tower and transmitted to the next tower by a second antenna. The other pair of antennæ is for traffic in the opposite direction. The average distance between towers is about 30 miles. The system operates in the 7-cm range and is capable of transmitting hundreds of telephone messages and several television programmes.

An interesting variant of the obstacle lens has been developed in France by Simon;[34] a microwave link system employing the new lens is in operation between Paris and Lille. This lens utilizes circular perforations in metal plates, and thus represents the "Babinet analogue" of the circular-disc obstacle medium. Some very good practical results have been obtained with a lens-mirror combination employing this principle, which has much to recommend it from the point of view of easy construction.

Various other lens configurations have been proposed, including a Lüneburg lens[25] employing artificial dielectrics. One version would employ the rodded medium developed at R.R.D.E.,[3,4] which appears to be capable of a very good match to free space.

In conclusion, it should be again emphasized that the utilization of obstacle-type artificial dielectrics in the construction of antenna lenses represents only one aspect of a rapidly expanding field, which includes such microwave applications as attenuators, dielectric-filled waveguides, end-fire antennæ, and non-reflecting walls.

8. Acknowledgments

This paper is based on an investigation sponsored in part by the U.S. Air Force.[37] The aid of Dr. F. J. Beck, of Yale University, in this investigation is also gratefully acknowledged.

9. Bibliography

1. Bouwkamp, C. J., "On the diffraction of electromagnetic waves by small circular disks and holes." *Philips Res. Rep.*, **5**, 1950, pp. 401-422.
2. Brown, J., "The design of metallic delay dielectrics." *Proc. Instn. El. Eng.*, **97**, Pt. III, 1950, pp. 45-48.
3. Brown, J., and Maclese, A., *Preliminary report on the rodded lens.* Report No. 348, Radar Research and Development Establishment, Malvern (Worcs.), September 1949.
4. Brown, J., and Jones, S. S. D., "Microwave lenses." *Electronic Engng*, **22**, 1950, pp. 127-131, *et seq.*
5. Campbell, R. D., and Schooley, E., "Spanning the continent by radio relay." *Bell Tel. Mag.*, **29**, 1950, pp. 215-227.
6. Cohn, S. B., "Analysis of the metal-strip delay structure for microwave lenses." *J. Appl. Phys.*, **20**, 1949, pp. 257-262 and p. 1011.
7. Cohn, S. B., "Electrolytic-tank measurements for microwave media." *J. Appl. Phys.*, **21**, 1950, pp. 674-680.
8. Cohn, S. B., "Determination of aperture parameters by electrolytic-tank measurements." *Proc. Inst. Rad. Eng.* **39**, 1951, pp. 1416-1421.
9. Cohn, S. B., "The electric and magnetic constants of metallic delay media containing obstacles of arbitrary shape and thickness." *J. Appl. Phys.*, **22**, 1951, pp. 628-634.

10. Cohn, S. B., *Investigation of metallic delay lenses.* Reports under U.S. Army Signal Corps Contract No. W36-039-sc-38250, Sperry Gyroscope Company, Great Neck (New York), 1949-1951.

11. Corkum, R. W., *Isotropic artificial dielectric.* Report from U.S. Air Force Cambridge Research Laboratories, March 1951. Abstracted in *Proc. Inst. Rad. Eng.*, **39**, 1951, p. 308.

12. Debye, P., "Der Lichtdruck auf Kugeln von beliebigem Material." *Ann. Phys.* (4.F.), **30**, 1909, pp. 57-136.

13. Estrin, G., "The effective permeability of an array of thin conducting disks." *J. Appl. Phys.*, **21**, 1950, pp. 667-670.

14. Estrin, G., "The effects of anisotropy in a three-dimensional array of conducting disks." *Proc. Inst. Rad. Eng.*, **39**, 1951, pp. 821-826.

15. Howes, K. J., and Whitehead, E. A. N., *The refractive index of a dielectric loaded with thin metal strips.* Report from the Research Laboratories of Elliott Brothers (London) Ltd., Boreham Wood (Herts.), July 1949.

16. van de Hulst, H. C., "Optics of spherical particles." *Rech. Astronom. d'Utrecht*, **9**, Pt. I, 1946.

17. Jahnke, E., and Emde, F., "Tables of functions." New York : Dover Publications, 1945 (4th ed.).

18. Jones, S. S. D., and Brown, J., "Metallic delay lenses." *Nature*, **163**, 1949, pp. 324-325.

19. Kock, W. E., "Metal lens antennas." *Proc. Inst. Rad. Eng.*, **34**, 1946, pp. 828-836.

20. Kock, W. E., "Metallic delay lenses." *Bell Syst. Tech. J.*, **27**, 1948, pp. 58-82.

21. Kock, W. E., "Path-length microwave lenses." *Proc. Inst. Rad. Eng.*, **37**, 1949, pp. 852-855.

22. Lewin, L., "The electrical constants of a material loaded with spherical particles." *J. Instn. El. Eng.*, **94**, Pt. III, 1947, pp. 65-68.

23. Lengyel, B. A., "A note on reflection and transmission." *J. Appl. Phys.*, **22**, 1951, pp. 263-264.

24. Leitner, A., and Spence, R. D., "The oblate spheroidal wave functions." *J. Franklin Inst.*, **249**, 1950, pp. 299-321.

25. Lüneburg, R. K., *Mathematical theory of optics.* Providence (Rhode Island): Brown University, Lecture Notes, 1944 (p. 213).

26. Meixner, J., and Andrejewski, W., "Strenge Theorie der Beugung ebener elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden ebenen Schirm." *Ann. Phys.* (6.F.), **7**, 1950, pp. 157-168.

27. Mie, G., "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen." *Ann. Phys.* (4.F.), **25**, 1908, pp. 377-445.

28. Montgomery, C. G., "Technique of microwave measurements." M.I.T. Rad. Lab. Series, Vol. 11. New York: McGraw-Hill Book Company, 1947.

29. Montgomery, C. G., *et al.*, "Principles of microwave circuits." M.I.T. Rad. Lab. Series, Vol. 8. New York: McGraw-Hill Book Company, 1948.

30. Möglich, F., "Beugungserscheinungen an Körpern von ellipsoidischer Gestalt." *Ann. Phys.* (4.F.), **83**, 1927, pp. 609-734.

31. Page, L., "Magnetic moments at high frequencies." *Phys. Rev.*, **60**, 1941, pp. 675-684.

32. Pasternack, S., *Microwave lenses.* Final Report under U.S. Air Force Contract No. W28-099-ac-112, University of Pennsylvania, July 1947.

33. Rust, N. M., "The phase correction of horn radiators." *J. Instn. El. Eng.*, **93**, Pt. IIIA, 1946, pp. 50-51.

34. Simon, J. C., "Étude de la diffraction des écrans plans: application aux lentilles hertziennes." *Ann. Radioélectr.* **6**, 1951, No. 25, pp. 205-243.

35. Stuetzer, O. M., "Development of artificial microwave optics in Germany." *Proc. Inst. Rad. Eng.*, **38**, 1950, pp. 1053-1056.

36. Süsskind, C,. "Microwave lenses." *Wireless World*, **55**, 1949, pp. 370-372.

37. Süsskind, C., *Investigation of obstacle-type artificial dielectrics.* Final Report under U.S. Air Force Contract No. AF 19 (122)-270, Yale University, June 1951.

38. Trinks, W., "Zur Vielfachstreuung an kleinen Kugeln." *Ann. Phys.* (5.F.), **22**, 1935, pp. 561-590.

DISCUSSION

D. G. Kiely: I note with interest the considerable difference between the curves of bulk refractivity plotted against element spacing for a lattice of dielectric obstacles and for a lattice of metal obstacles. Is there also a marked difference between the dispersive properties of artificial dielectrics composed of metal and dielectric obstacles, and which type of obstacle gives most dispersion?

Are the recent theories of artificial dielectrics capable of predicting the refractive index of a metal-strip medium to a sufficient degree of accuracy to enable lenses and prisms to be successfully designed without experimental assistance, particularly at frequencies at which the medium is dispersive?

A non-stepped plano-convex lens made from a metal-strip medium exhibited excessively large side lobes close to the main beam in the electric plane radiation pattern. The radiation pattern in the magnetic plane of a lens made from the same medium showed no abnormal behaviour. When the first lens was replaced by a similar equivalent polystyrene lens, the abnormal side lobes in the electric plane were not present. From his experience, can Dr. Süsskind offer an explanation of this effect?

Lt.-Col. J. P. A. Martindale (Associate Member): Engineering opinion in Great Britain and in the United States has appeared for nearly two years to be moving away from the "metallic delay" type of microwave lens.

In this country, broadly speaking, difficulties were experienced with construction—in view chiefly of the low mechanical strength of the electrically suitable materials—and in the non-availability of such materials, leading to an interest mainly in rod lenses. In these lenses, an array of rods parallel to the electric vector of the incident radiation represents the loading of free space with inductive elements, just as the strip-type metallic delay lens may be considered to load a region of space with capacitive elements.

In the United States, again to generalize, the main dissatisfaction was with the expenditure of man-hours in the tedious assembly of metallic delay lenses, and development effort in that country next turned to the "path length" type of lens.

Are we to understand that a renewed interest is now being taken in the United States in the metallic delay type of lens, and if so, what has led to this interest?

AUTHOR'S REPLY TO DISCUSSION

In reply to Mr. Kiely, those dispersive properties of artificial dielectrics which depend on obstacle *size* are more pronounced with conducting particles than with dielectric particles, and I have heard of at least one case where dielectric particles are used in preference to metal ones for precisely that reason, viz., to avoid dispersion. There are no data, to my knowledge, on the dispersion of dielectric-obstacle media as a function of obstacle *separation*, but I should think that this type of dispersion will be essentially the same as for conducting obstacles.

The transmission-line theory for metal-strip media is quite accurate, and I believe that agreement of actual properties of finished configurations with the values predicted from theory to within 1 per cent. is not uncommon. However, comparatively little work has been done on applications which are deliberately constructed to have dispersive properties, although such applications were suggested by Kock in his original paper (*cf.* Ref. 20 above): and I am not sure whether the agreement with theory would be quite as good at the higher frequencies.

A detrimental effect of the metal-strip lens on the radiation pattern has been observed by several investigators (see, for instance, p. 186 of Ref. 2 above). It has been variously ascribed to reflexions inside the lens, and to the fact that the field within the lens may not be everywhere normally incident to the strip faces; but the effect has not yet been completely explained.

In reply to Col. Martindale, there has been continued interest in obstacle-type artificial dielectrics in the U.S., as witness the various investigations undertaken in recent years, notably that carried out at the Sperry Gyroscope Company (*cf.* Ref. 10 above). The objections enumerated above are quite valid, and considerable thought has been devoted to developing new methods of construction. There is also a good deal of interest in artificial dielectrics *per se*, with a view to developing applications other than lens antennæ.

Col. Martindale has also communicated to me some misgivings regarding the proposed applications in which the refractive index is varied spatially, pointing out that for the two-dimensional case at least, a Lüneburg lens was most readily constructed by the *metal-plate* or "waveguide" method described by Jones (*Proc.Instn.El.Eng.*, **97**, Pt. III, 1950, pp. 255-258). In reply, it should perhaps be mentioned that the above lens has the disadvantage of not being capable of a 360° scanning action (which, in a Lüneburg lens, should be possible by rotating the feed alone around the lens), although it lends itself very well to unidirectional pencil-beam applications. A *path-length* type of lens which would yield the 360° scan has been designed in America; it was described by Fine in a paper presented in New York (abstracted in *Proc.Inst.Rad.Eng.*, **39**, 1951, p. 308). This so-called "geodesic-analogue" lens has the disadvantage that a toroidal bend or conical skirt termination must be added all around the circumference to produce a beam in the desired horizontal plane. These matching and termination problems should ideally be absent in a Lüneburg lens employing obstacle-type artificial dielectrics; and, in addition, it should be feasible to produce a three-dimensional model by this method ultimately.

DYNAMIC MEASUREMENTS ON RECEIVING VALVES*

by

A. J. Heins†

A Paper presented at the Second Session of the 1951 Radio Convention on July 5th, at University College, London

SUMMARY

A brief description is given of three types of dynamic test equipment for receiving valves to illustrate the fact that it is often possible to design this type of equipment in such a way that it is direct reading in units of the result required, at the same time avoiding the necessity of adjusting several controls. With such equipment batches of valves can be tested quickly on certain dynamic properties.

1. Introduction

In the manufacture of radio valves it is necessary to subject all valves to a limited number of measurements at the end of the manufacturing process in order to maintain a constant product. A small percentage of the production is then tested thoroughly in order to ensure that characteristic data which are not directly measured in the production tests remain at the correct level. These measurements usually cover such items as $I_a - V_g$ characteristic, slope, internal resistance, capacitances. It is often useful to augment these normally static measurements with checks on some dynamic properties such as hum, noise, cross modulation, power output.

The equipment necessary for these tests has been described elsewhere and is familiar to the technician who is engaged on valve measurements. However, most of these equipments require the manipulation of many controls, the use of a slide rule, or additional measurements, to obtain the result. It is, therefore, impracticable to subject large numbers of valves to these complicated measurements, and valuable information about the trend of the production may be lost. This can be improved by designing the measuring equipment in such a way that the manipulation of controls is kept to the minimum and that it is direct reading in units of the result required.

As examples of this principle, three types of test equipment will be described—measurement of equivalent noise resistance, measurement of power output and distortion, and measurement of cross modulation.

* Manuscript received May 26th, 1951.
† Mullard Radio Valve Co., Ltd.
U.D.C. No. 621.317 : 621.385.002.

2. Measurement of Equivalent Noise Resistance

The irregular arrival of electrons at the anode of a valve causes fluctuations in its anode current which, after amplification and detection, result in noise from the loudspeaker. Thermal agitation in the tuned circuits, particularly in the first stage, also contributes to this noise, and in order to assess easily the relative importance of the various noise sources it is convenient to express the valve noise as an equivalent noise resistance in the grid circuit which would cause the same noise in the anode circuit as the irregular arrival of electrons. This can only be done when transit time effects are negligible. At frequencies of 100 Mc/s or higher where this condition is not fulfilled the noise of a valve is usually defined by a "noise factor." There is no simple relation between noise factor and equivalent noise resistance, but it has been observed that once the design of a valve type is settled, noise factor and noise resistance show similar variations so that for a rapid check on a batch of valves it is of value to measure the equivalent noise resistance

In the audio frequency region up to about 10 kc/s irregular emission from the cathode can also cause noise. This is known as "flicker effect" and can only be measured at the frequencies concerned; the equipment to be described here is not suitable for this purpose.

The value of the equivalent noise resistance can be determined either by calculation from the slope of the valve and the value of the noise component in the anode current which can be measured by means of an amplifier of known band width and gain, or by comparing the noise caused by the valve with noise from a saturated diode or from a known resistance in the grid circuit of the valve. These methods although quite straightforward involve manipulation of various controls and slide rule calculations; they are, therefore, not suitable for making a rapid check on a batch of valves. The principle of these methods is the measurement of the noise component in the anode current by means of an amplifier of adequate gain and a pass band somewhere between 100 kc/s and 1 Mc/s so as to be outside the regions where flicker effect may occur, or where transit time becomes noticeable. In order to obtain a direct indication of the equivalent noise resistance it is necessary to include in the measurement the value of the slope of the valve under test. This can be realized by fitting the amplifier with an automatic gain control which by means of a 1,000-c/s auxiliary signal automatically adjusts the gain relative to the reciprocal of the slope of the valve under test as shown in Fig. 1.

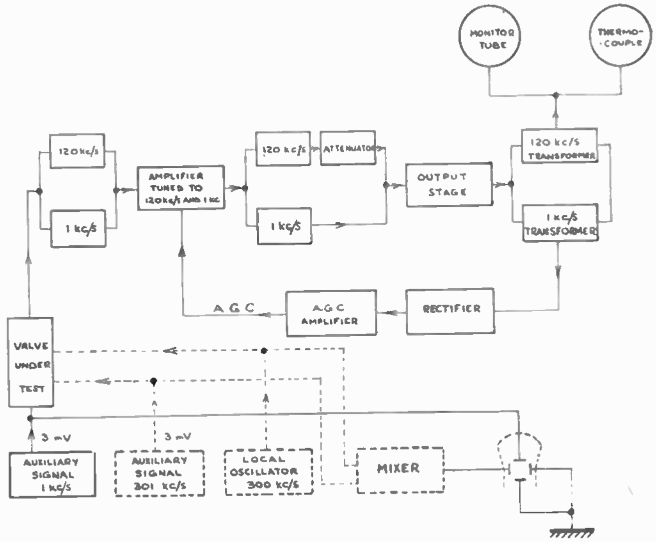


Fig. 1.—Block diagram of equipment for the direct measurement of equivalent noise resistance.

A small auxiliary signal of 3 mV at 1,000 c/s is fed to the grid of the valve under test; in the anode are connected a 120-kc/s transformer and a 1,000-c/s transformer both tuned to their respective frequencies. The amplitude of the auxiliary signal is sufficiently small not to cause any modulation effects and sufficiently large not to be affected by flicker effect. The noise of the valve under test passes through the 120-kc/s transformer into a multi-stage amplifier embodying variable-mu pentodes for automatic gain control. Between this amplifier and the output stage an attenuator is inserted in the 120-kc/s channel to act as a range switch for the indicating instrument. From the output stage the noise is fed through a 120-kc/s transformer to a thermo-couple instrument; the 1,000-c/s auxiliary signal, which has so far passed through the same valves as the noise, is rectified and compared with the voltage across a voltage reference tube. The d.c. voltage difference thus obtained is amplified and used to control the gain of the multi-stage amplifier. In this manner the gain of the 1,000-c/s channel from the grid of the valve under test to the output stage is kept constant within .1 per cent. for different values of slope of the valve under test between 25 μA/V and 25 mA/V. Since the same valves also amplify the noise it follows that the gain of the noise channel will also be kept constant, depending only on the position of the attenuator. The scale of the thermo-couple instrument can now be calibrated direct in units of equivalent noise resistance with the attenuator in the 120-kc/s channel acting as a range switch. In the actual equipment the attenuator has seven positions corresponding to one off position and full-scale

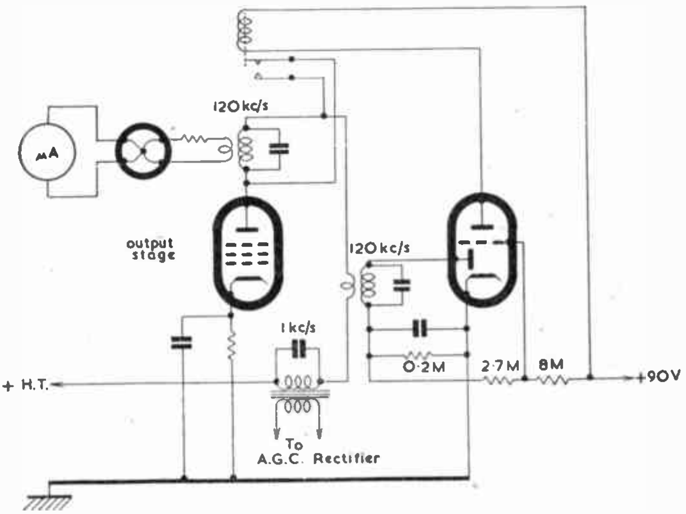


Fig. 2.—Circuit diagram of safety device to protect thermo-couple.

readings of 3, 10, 30, 100, 300 and 1,000 kΩ.

For the measurement of equivalent noise resistance of frequency changers the method just described needs some modification since it will then be necessary to include the conversion-conductance instead of the slope in the final result. This can be done by arranging a 1,000-c/s difference between the local oscillator for the frequency-changer and the auxiliary signal as indicated in broken lines in Fig. 1; the frequencies chosen are 300 and 301 kc/s. The correct frequency difference is monitored through a mixer and a monitor tube but since the 1,000-c/s channel is fairly wide small deviations from 1,000 c/s do not affect the result.

The instrument on which the equivalent noise resistance is read is a thermo-couple instrument and an automatic protection against overloads is necessary to avoid frequent burn-out of the thermo-couple. The method used is shown in Fig. 2.

The few turns of the primary of a second 120 kc/s transformer are connected in series with the output transformer; the tuned secondary feeds a diode detector and the resulting d.c. voltage is fed to the grid of a triode. A delay voltage is obtained through a 10MΩ resistance from a positive point. When an overload occurs the triode is cut off, thus releasing the relay in its anode circuit. The primary of the output transformer is short-circuited, thus preventing burn-out of the thermo-couple. The shorting of the output transformer results in a slight increase of the voltage on the diode; the action is therefore positive without chatter. When the overload disappears the triode becomes conducting again and the short circuit is removed. This device can be adjusted to operate on an overload of about 10 per cent.

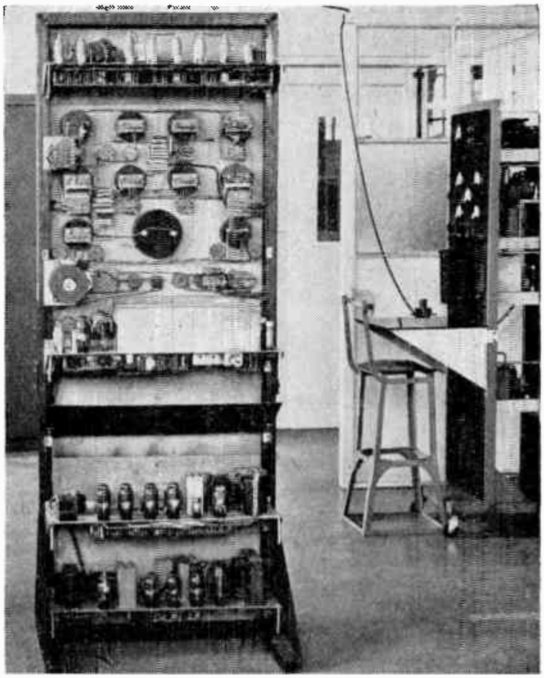


Fig. 3.—Rear view of equipment for the direct measurement of equivalent noise resistance.

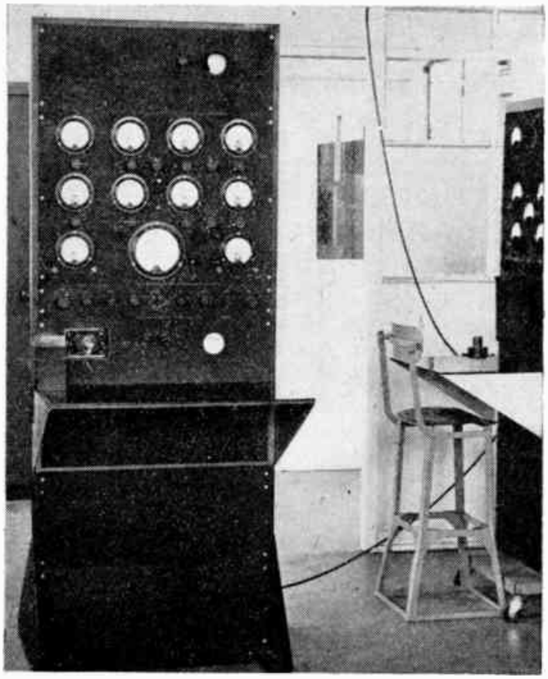


Fig. 4.—Front view of equipment for the direct measurement of equivalent noise resistance.

An equipment built along these lines and shown in the photographs (Figs. 3 and 4) has been in regular use for about six years; it has given no trouble, and, thanks to the protective device just described, still operates with its original thermo-couple. The electrode supply voltages are all obtained from stabilized supplies and once these are set for measurement on a certain valve type a batch of valves can be checked at the rate of about 150/hr.

3. Measurement of Power Output and Distortion

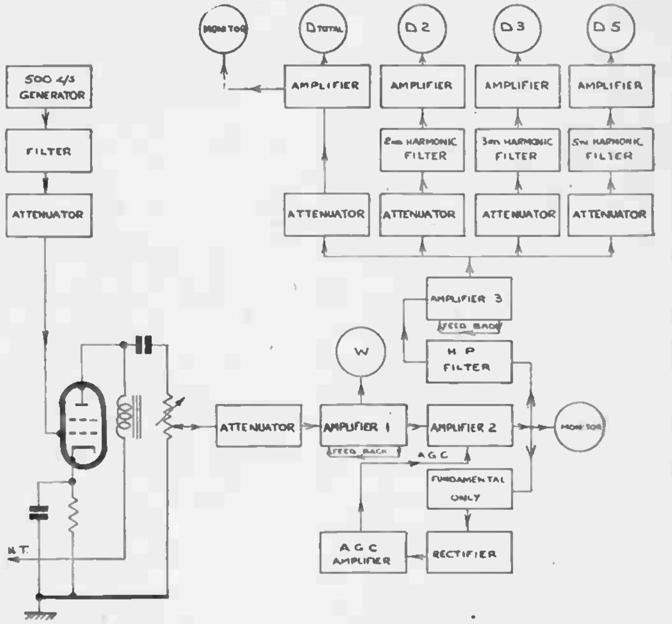


Fig. 5.—Block diagram of equipment for the measurement of power output and distortion.

Methods for the measurement of power output and distortion have been described by various authors. Although these methods can be relied upon to give accurate results when carried out with sufficient care, they are on the whole more or less laborious and not suitable for measuring a large batch of valves unless the method is modified to be direct reading in units of output and distortion. The block diagram of Fig. 5 shows one way of achieving this.

A pure sine wave of 500 c/s is fed to the signal grid of the valve under test which has a variable load resistance in its anode circuit; a variable tap on the load resistance is connected to an attenuator acting as a range switch for the output indicator, and this attenuator is followed by an amplifier with strong negative feed-back. The reason for this feed-back is two-fold: no extra distortion may be introduced by this amplifier since it precedes the distortion measurement, and secondly, with the strong feed-back, once the amplifier is calibrated it can be relied upon to maintain its calibration. A thermo-couple instrument is connected to the output of the amplifier. The variable tap on the load resistance is ganged with the main control knob which changes the total value of the load resistance and is so arranged that the value of the resistance below the tap is proportional to the square root of the total resistance, so that for a given power in the load resistance the voltage at the tap is the same for all values of load resistance. Bearing in mind that the deflection of a thermo-couple instrument is proportional to the square of the current passing through it, this load resistance tap arrangement results in a linear scale in watts on the instrument. In the actual equipment the wattmeter range switch has nine positions giving full-scale readings of 30 mW to 300 W.

The wattmeter amplifier (amplifier 1 in Fig. 5) is followed by an amplifier (No. 2 in Fig. 5) with a very rigorous automatic gain control which keeps the output level of the fundamental constant to within 1 per cent. for an input signal varying between approximately 20 and 200 mV, thus amply covering the difference between two successive stages of the wattmeter range switch. Since the distortion is measured after the signal has passed through this amplifier, the automatic control must be effected without introducing extra distortion on the one hand and without introducing inaccuracy due to the effect of valve noise on too low a signal level on the other. Preliminary measurements showed that if distortion factors down to 1 per cent. were to be measured the signal level at the input of this amplifier should be between about 20 and 200 mV. At this level the only distortion which might affect the accuracy of the measurement is the second harmonic; this has been avoided by the arrangement shown in Fig. 6. A straight pentode and a variable-mu pentode in cascade are connected in parallel with a variable-mu pentode and a straight pentode in cascade. The straight pentodes act as phase inverters and give no extra gain, the variable-mu pentodes are adjusted to give no gain at the maximum input to the amplifier of 200 mV. The second harmonic generated in the variable-mu valves is thus cancelled.

This arrangement is followed by two stages of amplification with negative feed-back and the output of this amplifier is connected: (*a*) to a monitor cathode-ray tube on which the wave-form of the distorted signal can be observed, (*b*) via a filter which passes the fundamental only to a rectifier and A.G.C. amplifier, and (*c*) to a high pass filter and following stages.

Since the automatic gain control in amplifier 2 keeps the level of the fundamental at its output constant, the direct indication of the percentage distortion now only requires straightforward filters and amplifiers as shown in Fig. 5. In order to avoid overloading in subsequent filters and amplifiers, the fundamental is thoroughly suppressed in a high pass filter, and after amplification the distortion is fed to four channels, thus giving on four separate instruments readings for total distortion, second-, third-, and fifth-harmonic. The wave-form of the distortion only can furthermore be observed on a monitor cathode-ray tube connected to the output of the total distortion amplifier. Each channel has a range switch giving full-scale deflection for 1, 3, 10 and 30 per cent. distortion. The residual distortion of the whole equipment is of the order 0·05 per cent.

After adjustment of the valve operating conditions it is only necessary to operate the input signal control; power output and distortion can then be read directly on the appropriate instruments. In this way, optimum operating conditions including load resistance can be established very rapidly or, alternatively, once these are known for a certain valve type, the performance of a batch of valves can easily and quickly be checked.

The equipment just described and shown in Fig. 7 has been in regular use for several years and has given no trouble.

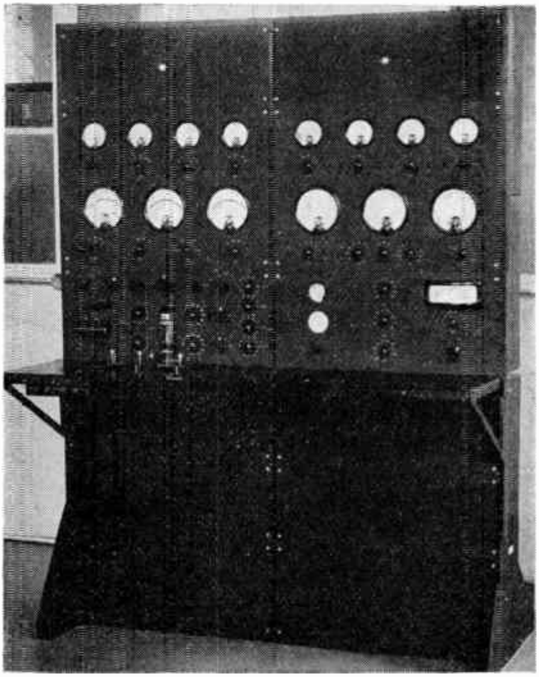


Fig. 7.—Front view of equipment for the measurement of power output and distortion.

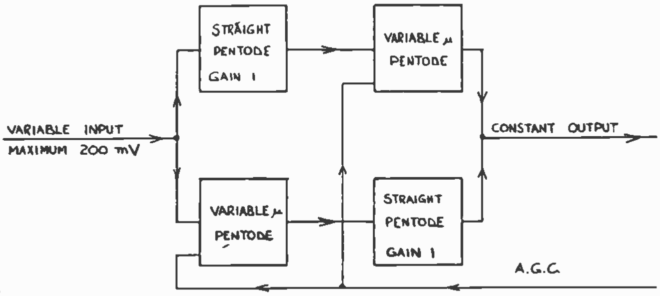


Fig. 6.—Block diagram of valve arrangement to avoid the introduction of second harmonic by gain control in amplifier 2.

4. Measurement of Cross Modulation

Cross modulation is the effect whereby the modulation of an interfering signal appears on the carrier of a desired station. The ratio of modulation depth caused by the interfering signal to its original modulation depth is called cross modulation factor, and for normal broadcast receivers a cross modulation factor of ½ to 1 per cent. is usually tolerated. For design calculations of radio receivers and for comparison of different variable-mu pentodes or frequency-changers it is convenient to represent the signal handling capacity in a graph of the signal which causes 1 per cent. cross modulation as a function of slope or conversion conductance.

An equipment designed to take these curves rapidly is shown in block diagram in Fig. 8.

A 100-mV unmodulated carrier of 120 kc/s and a modulated interfering signal at 1·5 Mc/s are fed to the signal grid of the valve under test.

The frequencies are chosen so far apart to ensure that cross modulations will occur only in the valve under test and not in the following amplifiers. The valve under test is followed by an amplifier tuned to 120 kc/s which in turn is followed by an attenuator acting as range switch for the slope indication. From this attenuator the signal is fed to two channels, one incorporating an amplifier with constant gain and tuned to 120 kc/s terminating in a diode voltmeter calibrated to read the slope of the valve under test. The other channel contains an amplifier also tuned to 120 kc/s with a very rigorous automatic gain control which keeps the signal at the following detector constant within 1 per cent. After detection, the modulation which originates from the interfering signal is amplified and rectified. The d.c. voltage thus obtained is an indication of the modulation depth due to cross modulation. After comparison with the voltage drop across a voltage reference tube the d.c. voltage is amplified and used for automatic control of the amplitude of the interfering signal. In this way the interference is automatically adjusted to the correct level for a cross modulation factor of 1 per cent.

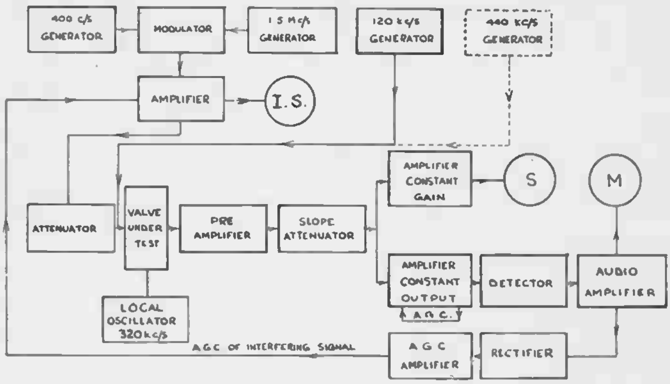


Fig. 8.—Block diagram of equipment for the measurement of cross modulation.

The d.c. voltages for the electrodes of the valve under test are obtained from stabilized voltage supplies, and after these have been adjusted to their correct values the input voltage curve for a cross modulation factor of 1 per cent. can be plotted by varying the grid bias only, and reading on the appropriate instruments the values for slope, anode current and interfering signal.

For the measurement of frequency-changers, the unmodulated 120-kc/s signal is replaced by a 440-kc/s signal, which with a local oscillator frequency at 320 kc/s gives a beat frequency of 120 kc/s, so that the rest of the equipment remains unchanged.

5. Conclusion

In the introduction it was recommended that the normal production and sample testing of receiving valves be augmented by a check on some of the dynamic properties. The objection to this is that many of the dynamic measurements which might give useful information on the quality of the product require too much time. The equipments described in this paper illustrate that it is often possible to design the apparatus in such a way that the operator needs to adjust no controls, or only very few, and that the result can be read directly on a meter.

6. Acknowledgments

As mentioned earlier these equipments have been in regular use for several years without giving trouble, and the author wishes to acknowledge that this is largely due to the assistance of Mr. F. Amstelveen who has so admirably planned and supervised the building of these equipments.