The Journal of the

BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 24 JULY 1962 NUMBER 1

INTERNATIONAL ACHIEVEMENT

THE TELSTAR experimental communications satellite which was launched on 10th July last has provided proof of the feasibility of using artificial earth satellites as radio relay stations in long distance television, telephony and telegraphy. Excellent results were obtained for transmission between the ground station at Andover, Maine, U.S.A., and Goonhilly Downs, Cornwall, England, and Lannion in Brittany, France, via the TELSTAR satellite which was developed by the American Bell Telephone Laboratories.

The satellite orbit is elliptical with a period of approximately 2 hours 40 minutes inclined at about 50 deg to the Equator and with its apogee at about 3500 miles and perigee about 600 miles. The maximum period of mutual visibility of the satellite from the two sides of the Atlantic Ocean is only 35 minutes and hence for any future practical system a number of satellites will be needed. The experiment has however shown that the satellite can be successfully tracked and that the transmitter/receiver circuits it carries can handle the broadband signals represented by 600 channel telephone transmissions and colour television. Transmission to the satellite is at 6390 Mc/s and after considerable amplification (over 90 dB) first at an i.f. of 90 Mc/s and then at r.f. by travelling-wave tube, the signal is re-transmitted to ground at 4170 Mc/s. The system bandwidth is 50 Mc/s.

The TELSTAR project has called for close international co-operation between American, British and French engineers. British Post Office engineers designed and built the Goonhilly Downs Station in the short space of less than a year—in itself a commendable achievement. British industry has supported this effort to the full, notable among the participants being the Mullard Research Laboratories who designed, constructed and installed the travelling wave maser for reception in only six months.

While telephone traffic will obviously be the main contributor to the future use of satellite systems, the more spectacular application of international television exchanges will also feature more and more frequently although time differentials will limit the scope of live programme transmission. Nevertheless the existence of satellite facilities will stimulate new ventures in promoting international exchanges of news, views and entertainment of varied nature. The Institution has suggested† that British television receivers should be designed to reproduce pictures on American standards, and if this proposal were adopted, programmes could then be fed into the British distribution network without standards conversion.

A second experimental communications satellite similar to TELSTAR and to be known as RELAY will be launched later this year. Although proposals for systems of satellites have been put forward from several quarters (for instance, in papers read at last year's Brit.I.R.E. Convention on "Radio Techniques and Space Research"), firm plans for the initiation of a full-scale scheme have so far not been announced. There is world wide interest in the possibilities of satellite communications and the recent Commonwealth Communications Conference in London (as reported in the June *Journal*) recognized the part which such systems can play in linking the Commonwealth countries.

---

† See "The Report of the Committee on Broadcasting" on pages 4 and 5 of this *Journal*.

The Report of the Committee on Broadcasting 1960

The Committee on Broadcasting presented its report (Cmd. 1753) to the Postmaster General on 5th June, 1962. It is gratifying to record that many of the Institution's recommendations found favour with the Committee.*

Two sections of the Institution's survey dealing with a National Communications Authority and with wired systems appear to have been misunderstood by the Committee. This supplementary report is therefore issued to clarify the Institution's recommendations and also to comment further on standards and systems for colour television transmission.

National Communications Authority (N.C.A.)

The Institution recommended in Section 11 of its Survey the appointment of a single National Communications Authority to replace the technical planning activity of several existing bodies. Such an Authority would make regulations on all matters relating to system standards, frequency allocation, interference between services, service areas, wired broadcasting licences, etc., and ensure compliance with agreed regulations.

It appears from paragraph 424 of the Committee's Report that the above recommendation was misinterpreted as including the control of broadcasting *content* whereas the recommendation referred only to the control of *technical* standards and *systems* of broadcasting.

The main reason for suggesting the formation of a National Communications Authority was to ensure proper utilization and allocation of all radio frequencies. In Great Britain, the allocation of frequencies is entirely under the control of the Postmaster General. The Post Office is a considerable user of radio frequencies, and is likely in the future to be an even greater user. The advantage, therefore, of setting up an Authority is that there would be better opportunity for all users, especially the Defence Services, to explain their claims for frequency allocations to an independent tribunal. The Institution further recommended that a National Broadcasting Systems Committee should be appointed which would act as an advisory body to the main authority on such technical problems as systems of colour and monochrome television, stereophonic sound broadcasting, etc.

Future development of communications, e.g., the possibilities of using communication satellites, emphasizes the immediate and future problem of adequately using the already crowded frequency spectrum. There is also the need to have a strong national authority to negotiate international agreements. Near chaotic conditions already prevail in the world-wide use of radio frequencies, and the future development of communications may well be threatened by not safeguarding frequencies essential to the national interest.

Wired Systems as an Alternative to U.H.F. Transmissions

When the Committee on Broadcasting was appointed, its terms of reference were:

> "To consider the future of the broadcasting services in the United Kingdom, *the dissemination by wire* of broadcasting and other programmes . . ."†

Consequently, in the Institution's survey (Section 4), consideration was given to the provision of wired systems as *an alternative* to U.H.F. transmissions and as part of a national distribution service. Referring to this submission, the Committee stated (para. 730):

> "It seems to us that the case for the use of transmission by wire as the *primary* means of transmission would need to be convincingly demonstrated. We are informed that such a development has *never* been considered by T.A.C. nor referred to the Postmaster General."†

This comment is especially surprising in view of the inherent experience and acknowledged achievement of the Post Office Engineering Department in the development of communication by wire.

The British Post Office has gained a wealth of experience as a licensing authority, by specifying standards of technical performance for wired relay concessionaires, quite apart from its own work in the distribution of telephone and television traffic. In the latter connection, the Post Office supplies facilities for closed circuit television, and provides links for both the B.B.C. and I.T.A.

* "Radio and Television Broadcasting in Great Britain." A survey prepared by the British Institution of Radio Engineers. Submitted to the Committee on Broadcasting on 12th May, 1961, and published in *J.Brit.I.R.E.*, **21**, pp. 379–387, May 1961.

† *For the purpose of emphasis in these comments certain passages of the Committee's Report have been printed in italics.*

In order to pay its way for future development, the Post Office is well alive to the technical possibilities of increasing the facilities it can offer for telephone* and other services through a National Communications Network. Unfortunately, the bandwidth capabilities of the present-day telephone circuit are not adequate to provide the public with all the services it may reasonably want in the foreseeable future, particularly in view of the scientific and technical progress which is now being made. It would seem that the present opportunity is one which should not be missed for providing facilities which would at the same time relieve congestion on the ether. The Institution therefore recommended in Section 4 of its Survey the establishment of a national closed circuit communications network which could now use wide-band coaxial cable or microwave links, and which in future may use waveguide techniques, or any suitable combination of these as circumstances *and developments* permit. The technical advantages of wire transmission are acknowledged in paragraph 729 of the Committee's Report.

Such a network already partly exists in the form of existing G.P.O. trunk routes. Indeed, before programme material can be radiated by the B.B.C. and I.T.A. transmitters, it must first be carried over Post Office coaxial cable or microwave links from studios, etc., to the individual transmitters. In ten years the Post Office network has grown from 1000 to 6000 channel miles. By 1964 the main television networks are planned to be expanded to 7100 channel miles.

Some of the additional communication services which could be provided by a national closed-circuit communications network are:

1. Telephone visual service.
2. Subscription television programmes (see recommendation 10 of the Institution's report).
3. Vision links between schools, colleges and universities for educational purposes.
4. Vision links for large screen theatre television.
5. Vision links between medical centres which could, for example, provide immediate consultation with a central hospital.
6. Slow-scan (narrow band) television for special services and facsimile reproduction of documents. (These facilities could be of great value to medical, civil and military services as well as to commerce.)
7. Local and national sound radio services, information and weather forecasts.

New Television Standards

Although the Institution recommended a change to the 625 line standard and this has been accepted by the Committee on Broadcasting and the Government in its White Paper,† the Institution wishes to stress the advantages of an increased picture frequency. This was dealt with in Section 5 of the Institution's survey. It was recommended that, for the future, plans should be made for the transmission of either 625 lines 25 pictures per second or 525 lines 30 pictures per second. If manufacturers were encouraged to make relatively simple modifications to television receivers so that they could accommodate transmissions on both standards, the advantages of this dual standard operation could be utilized at a later date.

International adoption of a single standard for television transmission has still a long way to go, but with a system capable of operating on either of the two main world standards, the United Kingdom would be well placed to take advantage of the benefits of both.

Furthermore there is a very noticeable reduction in flicker with the higher picture frequency, particularly at higher brightness levels.

Colour Television System of Transmission

It has been noted that neither the Government nor the Pilkington Committee has recommended a particular system of colour transmission. It is the Institution's view (Section 8) that the choice of colour systems should be decided on technical merit after suitable technical studies and investigation of the possibilities of international agreement. If the Government eventually sets up the National Communications Authority with its technical committee the National Broadcasting Systems Committee, it is this latter committee which would be capable of a realistic decision on such questions as the system of colour television to be used. At the present time there is no committee representative of the broadcasting authorities, programme originators and the receiver and transmitter designers, who can reach a decision on this matter or promote specific studies of these problems.

* The Annual Reports of the Telecommunication Engineering and Manufacturing Association for 1960–61 and 1961–62 show that the United Kingdom lags behind other major industrial countries in the expansion of internal communication services.

† Cmd. 1770.

INSTITUTION NOTICES

Birthday Honours List

The announcement on page 414 of the June *Journal* incorrectly stated that Squadron Leader W. R. F. Cooney had been appointed an Ordinary Officer of the Military Division of the Most Excellent Order of the British Empire. Squadron Leader Cooney's appointment was to be an *Ordinary Member* of the Order.

Award of the First Mountbatten Research Studentship

The first Mountbatten Research Studentship has been awarded to Mr. Robert D. Pringle, B.Sc., to enable him to carry out research on masers in the Department of Physics at the University of Edinburgh. The establishment of the Mountbatten Research Studentships, which have a value of £500, was announced in the April 1962 *Journal.*

The 1963 Convention Committee

The following members have been appointed by the Council to organize the 1963 Convention. The Convention will be held at the University of Southampton from 16th-20th April, 1963, and will be related to National Productivity Year.

Chairmen: J. L. Thompson (Member)
Vice-Chairman: A. St. Johnston, B.Sc. (Member) and
R. A. Bent (Member)
W. A. Gambling, Ph.D. (Associate Member)
J. R. Halsall, Dip.El. (Associate Member)
P. Huggins (Member)
M. James, Dip.El. (Member)
K. G. Nichols, M.Sc. (Associate Member)
W. Renwick, M.A., B.Sc. (Member)
T. S. Worthington (Associate Member)

The Committee is planning sessions under the following headings:

Measurement and Sensing Devices. (Papers are invited on transducers for quantitative measurements (e.g. pressure, temperature, position, optical factors etc.), quality sensing, objective and subjective aspects.)

Information Transmission and Communication. (Papers are invited on, for instance, digital and analogue data transmission and telemetering.)

Control and Information Processing. (Papers are invited, for example, on mathematical and logical techniques, storage, function generators, controllers, data loggers and computers.)

Output Devices and Final Control Elements. (Papers are invited on actuators and displays.)

Industrial Applications of Electronic Systems. (Papers are invited on, for instance, machine tool and positional control; inspection and non-destructive testing; process control systems; electronic methods in production; production control systems.)

Members are invited to submit papers for consideration for inclusion in the Convention. Offers should be accompanied initially by a synopsis of 200 words and this should be sent to the Institution as soon as possible.

North-Western Section Visit

The Committee of the North Western Section has arranged for members of the Section to visit the A.B.C. Studios, Didsbury, Manchester on Monday, 3rd September, at 7 p.m. A lecture accompanied by demonstrations will be given on "The SECAM System of Colour Television". Accommodation is limited and admission will be *by ticket only* obtainable from the Honorary Local Secretary of the Section, Mr. F. A. Mitchell, 12 Hillcrest Road, Offerton, Stockport, Cheshire.

Membership of the Group Committees

It is the intention of the Council that the membership of the Specialized Group Committees should be fully representative of the interests of members of these Groups. To this end, nominations of Corporate Members to serve on the following Group Committees are invited:

Computer Group Committee
Electro-Acoustics Group Committee
Medical and Biological Electronics Group Committee
Radar and Navigational Aids Group Committee
Television Group Committee

All nominations should be supported by two Corporate Members and be accompanied by the nominee's agreement to serve if appointed. Nominations must be received by the Secretary not later than 1st September 1962.

Telefis Eireann

The Director of the European Broadcasting Union has requested that a statement in the above news item published in the June 1962 *Journal* (page 430) should be clarified. The assignment of frequencies to Telefis Eireann for television broadcasting was made by the *European Broadcasting Conference* in Stockholm in 1961. The European Broadcasting Union does not have regulatory powers in the matter of frequency assignments.

Correction

The following amendments should be made to illustrations in the paper "Transistor Video Amplifier and Line Time-base Synchronization Circuits for Television Receivers", published in the June 1962 issue of the *Journal*,

Page 447, Fig. 3. The resistor from the base of the final transistor should not be connected to h.t. (140 volts) but to the lower end of the collector load resistor R_L.

Page 451, Fig. 7. The short circuit connection across the diode D2 should be deleted.

U.D.C. 621.397.332: 621.382.3

Television Line Time-Base Output Stages Using Transistors

By

K. E. MARTIN†

Based on a contribution to the Television Group Discussion Meeting on "Transistorized Television Receivers" in London on 6th April 1961.

Summary: The paper discusses the peak VA products to be handled by the line output transistor and efficiency diode for 70 deg., 90 deg. and 110 deg. picture tubes with "wide" and "narrow" necks. Various aspects of the circuit design are described and a complete design given of the output stage for a 90 deg. picture tube operating at 12 kV.

List of Symbols

F	ratio of peak voltage to average voltage
F_D	deflection factor
I_{ht}	average h.t. current
$i_{T(pk-pk)}$	peak-to-peak current in transformer primary
$i_{y(pk-pk)}$	peak-to-peak current in scanning coil
k	transformer coupling factor
L_L	leakage inductance
L_p	transformer primary inductance
L_s	transformer secondary inductance
L_y	scanning coil inductance
n	$\frac{n_{sec}}{n_{pri}}$
n_1	$\frac{n_{overwind}}{n_{pri}}$
n_{pri}	number of turns of primary winding
n_{sec}	number of turns of secondary winding
t_f	time of flyback period
t_s	time of scan period
t_t	time of total line period
V_b	boost voltage
$V_{ce(pk)}$	peak voltage between collector and emitter
V_{df}	forward voltage across diode
V_{ht}	h.t. potential
$v_{pk(pos)}$	positive peak voltage
V_s	transistor saturation voltage
V_{sr}	ripple voltage superimposed on h.t. line
$V_{T(pk-pk)}$	peak-to-peak voltage across transformer primary
V_w	voltage drop across circuit wiring
$V_{y(av)}$	voltage across the scanning coils during the scanning period
$v_{y(pk)}$	peak voltage across the scanning coils during the flyback period
V_{yr}	voltage across scanning coil resistance

1. Introduction

Transistor television receivers fall into two main categories, portable receivers and "full performance" mains receivers using 110 degree picture tubes operated at 16 to 18 kV. Even a portable receiver would be operated from the mains for most of its life, but it would have facilities for working from its own batteries or from an external battery, e.g. a motor car battery. A mains receiver would be designed mainly for operation from the mains, but it may have facilities for battery operation, possibly with lower performance.

If the receiver is intended for portable operation, then consideration must be given to the size and weight of the receiver and in particular to the picture tube. This type of receiver would most probably use a smaller tube than the "full performance" receiver and since it must be capable of operating from, say, a lead-acid car battery, the h.t. line will be determined by the battery voltage and not necessarily by the circuit requirements. If the "full performance" receiver has portable facilities, then again the h.t. line will be determined by the battery voltage. If, however, the receiver is intended for only mains operation, it could be that the optimum h.t. line would be in excess of that given from a lead-acid battery.

For a receiver operating from batteries, efficiency of the line time-base is a very important factor, since the power taken by the time-base will mainly determine the operating time of the receiver for a given battery size. If the receiver is designed for only mains operation, then efficiency is not quite so important, although of course the greater the circuit efficiency the smaller and cheaper the line output transistor becomes.

In this paper line time-bases for portable receivers are considered where the receiver must be capable of

† Mullard Research Laboratories, Redhill, Surrey.

operating from a lead-acid battery with a nominal voltage of 12·6 V and internal batteries of the same voltage, so that circuit efficiency is of major importance.

2. Peak VA Products

2.1. *Peak VA Product in the Scanning Coils*

In the Blumlein type of energy recovery circuit two switching devices are used; each conducts during the scanning period but not necessarily together and are open circuited during the flyback period. A large peak voltage will occur across them during the flyback period and each will pass a sawtooth pulse of current during the scanning period. It is convenient to express the requirements of the switches, in this case the transistor and efficiency diode, in terms of the peak value of the volt-ampere product to be handled by each device. It is therefore necessary to know the peak VA product needed by the scanning coils to give the required scan for a particular application.

Now
$$V_{y(\mathrm{pk})} = FV_{y(\mathrm{av})} \qquad \ldots\ldots(1)$$

If it is assumed that the peak voltage waveform is a half sine wave and that 3rd harmonic tuning is used then

$$F = \frac{0{\cdot}85\pi t_t}{2t_f} \qquad \ldots\ldots(2)$$

Therefore scanning-coil peak VA product

$$= \frac{FL_y(i_{y(\mathrm{pk-pk})})^2}{t_s}$$

$$= \frac{0{\cdot}85\pi L_y(i_{y(\mathrm{pk-pk})})^2 t_t}{2t_s t_f}$$

If a flyback period of 22% of the line period is assumed, then scanning-coil peak VA product

$$= 78 L_y(i_{y(\mathrm{pk-pk})})^2 \qquad \ldots\ldots(3)$$

where L_y is in mH. Figure 1 shows the required peak VA product of the scanning coils for various picture-tube scanning angles and neck diameters plotted against e.h.t. potential.

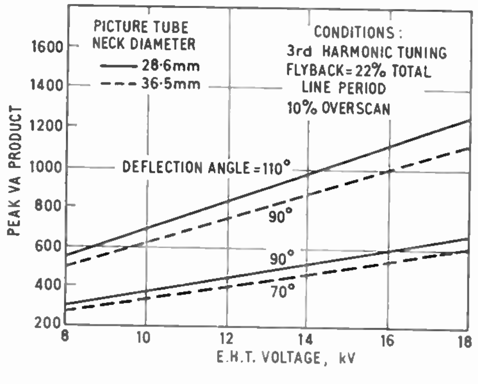


Fig. 1. Scanning coil peak VA vs. e.h.t. voltage.

2.2. *Peak VA Product of the Output Devices*

If the output transformer had infinite inductance and a coupling coefficient of unity, then the output devices would have to handle between them the same peak VA product as the scanning coils. However, because of the losses in the transformer, the peak VA product handled by the devices will be increased by about 10%.

Because of resistance in the circuit, the scan will be non-linear and if some form of non-linearity correction is used, then the transistor peak VA product will be increased by about a further 20%. This leads to the peak VA product requirement of the devices being 30% greater than that of the scanning coils.

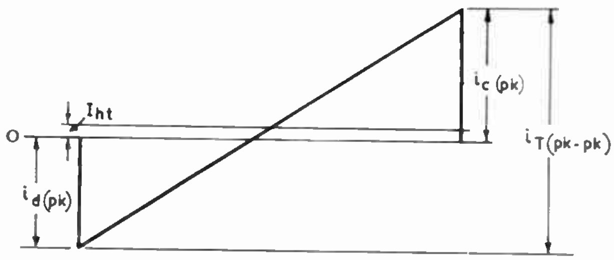


Fig. 2. Currents in output stage.

If a line time-base using a transistor as the output device and a diode for energy recovery is now considered, then the current flowing in the transistor and diode is as shown in Fig. 2. Because of the losses in the circuit, the peak value of the transistor current $i_{c(\mathrm{pk})}$ will be greater than that of the diode $i_{d(\mathrm{pk})}$ and in practice

$$i_{c(\mathrm{pk})} \simeq \frac{i_{d(\mathrm{pk})}}{0{\cdot}8} \qquad \ldots\ldots(4)$$

From Fig. 2

$$i_{T(\mathrm{pk-pk})} = i_{c(\mathrm{pk})} + i_{d(\mathrm{pk})} \qquad \ldots\ldots(5)$$
$$= i_{c(\mathrm{pk})} + 0{\cdot}8\, i_{c(\mathrm{pk})}$$

Therefore,

$$i_{c(\mathrm{pk})} = \frac{i_{T(\mathrm{pk-pk})}}{1{\cdot}8} \qquad \ldots\ldots(6)$$

If allowance is made for the losses of the transformer and the non-linearity correction, then transistor peak VA product

$$= \frac{1{\cdot}3}{1{\cdot}8} \times \text{scanning coil peak VA product.}$$

In Fig. 3 the transistor peak VA product is plotted against e.h.t. potential for various picture tube scanning angles.

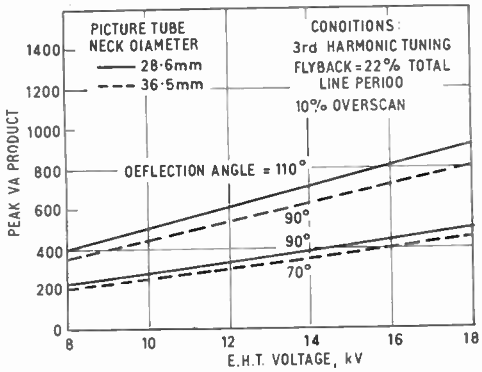


Fig. 3. Output transistor peak VA vs. e.h.t. voltage.

From equations (4) and (5)

$$i_{d(\mathrm{pk})} = \frac{i_{T(\mathrm{pk-pk})}}{2{\cdot}25}$$

Therefore

efficiency diode VA product

$$= \frac{1{\cdot}3}{2{\cdot}25} \times \text{scanning coil peak VA product.}$$

3. Choice of Picture Tube

For a portable receiver a choice of picture tube must be made bearing in mind,

(*a*) scanning angle

(*b*) e.h.t.

(*c*) weight of tube.

A narrow angle tube is an advantage for keeping the power consumption of the time-bases to low values, but the length of tube may defeat the object of portability. The e.h.t. must be high enough to give a bright, sharp picture and yet a compromise may have to be made between brightness and power taken by the time-base. It is therefore desirable to choose the lowest e.h.t. consistent with acceptable picture brightness. It will be seen from Fig. 1 that for the 90 deg. tubes the VA required for the narrow neck (28·6 mm) is only 60% of that for the wide (36·5 mm) neck. Although the 70 deg. tube even with a wide neck requires less power than the narrow neck 90 deg. tube, and a narrow neck 70 deg. tube would require even less, 70 deg. tubes are long and bulky and would be unsuitable for a portable receiver.

Table 1 shows the comparison between a 110 deg. tube operated at 16 kV and a narrow neck 90 deg. tube working at 12 kV, which is about the lowest acceptable e.h.t.

From the table it will be seen that the 90 deg. narrow neck tube at 12 kV requires only 40% the VA of the "full performance" 110 deg. picture tube at 16 kV.

Table 1

Deflection Angle	E.H.T. (kV)	Peak VA Product Transistor	Peak VA Product Diode
90°	12	325	260
110°	16	808	646

Neck diameter = 28·6 mm.

4. Circuit Arraugements

The peak VA product for the transistor and efficiency diode are now known for any particular application, but the actual values of peak voltages and peak currents will depend on the type of circuit and the h.t. potential used.

The two basic forms of energy-recovery circuit are shown in Fig. 4. The circuit of Fig. 4 (*a*) uses a series or booster diode to charge C1 to a potential higher than that of the h.t. line. This is the familiar circuit used for line output stages using valves. The advantage of this type of circuit is the boosted h.t. line, and the output devices are operating from a high voltage with low current which is especially suitable for valves.

Figure 4 (*b*) shows the parallel diode circuit where current is fed back to the power supply during the first part of the scan stroke (diode condition period). This is essentially a lower voltage, higher current method of operation.

Figures 4 (*c*) and 4 (*d*) show the voltage waveforms appearing across the transformers in the two circuits. It is assumed in both circuits that, during the scanning period, the mean value of the collector-to-emitter voltage is zero which, in fact, is a very close approximation. Then

$$v_{T(\mathrm{pk})} = v_{ce(\mathrm{pk})}$$
$$= FV_{ht} \text{ in Fig. 4 } (b)$$
$$= F(V_{ht} + V_b) \text{ in Fig. 4 } (a)$$

If the flyback period is 22% of the total line period, then from eqn. (2), F is 6·06. If the time-base is operating from a lead-acid battery with a nominal h.t. line of 12·6 V, then for Fig. 4 (*b*),

$$v_{ce(\mathrm{pk})} = 12{\cdot}6 \times 6{\cdot}06 = 77\ \mathrm{V}$$

If the circuit of Fig. 4 (*a*) is designed for the same diode conduction time and h.t. potential, then

$$v_{ce(\mathrm{pk})} = 2FV_{ht} = 154\ \mathrm{V}$$

Therefore, the series diode energy-recovery circuit of Fig. 4 (*a*) imposes twice the peak voltage across the transistor as the circuit of Fig. 4 (*b*), but only half the peak current. The diodes in both circuits will have the same operating conditions for peak current and voltage.

Transistors which are suitable for line output stages have v_{ce} ratings of 120 V. Some transistors are allowed 160 V between collector and emitter, but these are usually expensive. These ratings are absolute maximum values which must not be exceeded under the worst probable operating conditions taking into account production spreads in components, variations in h.t. potential, and circuit operational variations (picture tube beam current and "out of synch" condition). The circuit must therefore be designed such that under nominal conditions the v_{ce} during the flyback period is appreciably less than the absolute maximum rating of the transistor. With a transistor rating of 120 V the circuit in Fig. 4 (*a*) is therefore unsuitable for operation from a 12·6 V h.t. line with 22% flyback since $v_{ce(pk)}$ is 154 V. The circuit of Fig. 4 (*b*) giving a value of $v_{ce(pk)}$ of 77 V is acceptable for the 12·6 V h.t. line.

If the h.t. line was say 6·3 V, then circuit Fig. 4 (*a*) would be used, since $v_{ce(pk)}$ would then be 77 V. So circuit Fig. 4 (*a*) is usually only used if the h.t. line is low, and when boosting it $v_{ce(pk)}$ will still be within the transistor rating.

It is interesting to see that $v_{ce(pk)}$ depends only on the ratio t_t/t_f and V_{ht} since

$$v_{ce(pk)} = \frac{0{\cdot}85\pi V_{ht} t_t}{2t_f} \qquad \ldots\ldots(7)$$

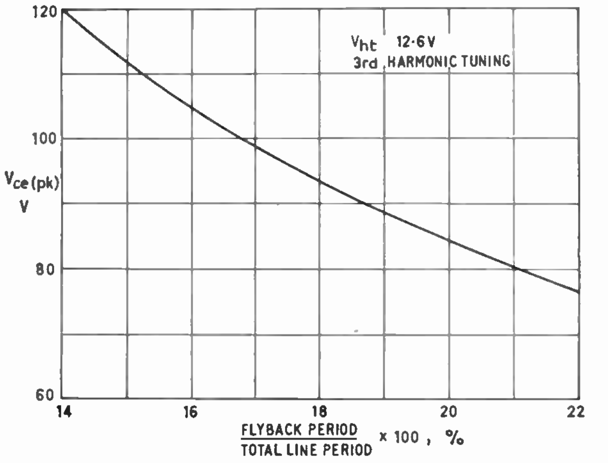


Fig. 5. Variation of peak collector to emitter voltage with flyback period.

This is true for any scanning power or television system, so if t_t/t_f is fixed, $v_{ce(pk)}$ will be directly proportional to V_{ht}.

5. Flyback Time

In the preceding sections, a flyback period of 22% of the total line period has been assumed. Some justification for this should be given. Figure 5 shows $v_{ce(pk)}$ plotted against percentage flyback period for a 12·6 V h.t. potential. If t_f was made less than 21%

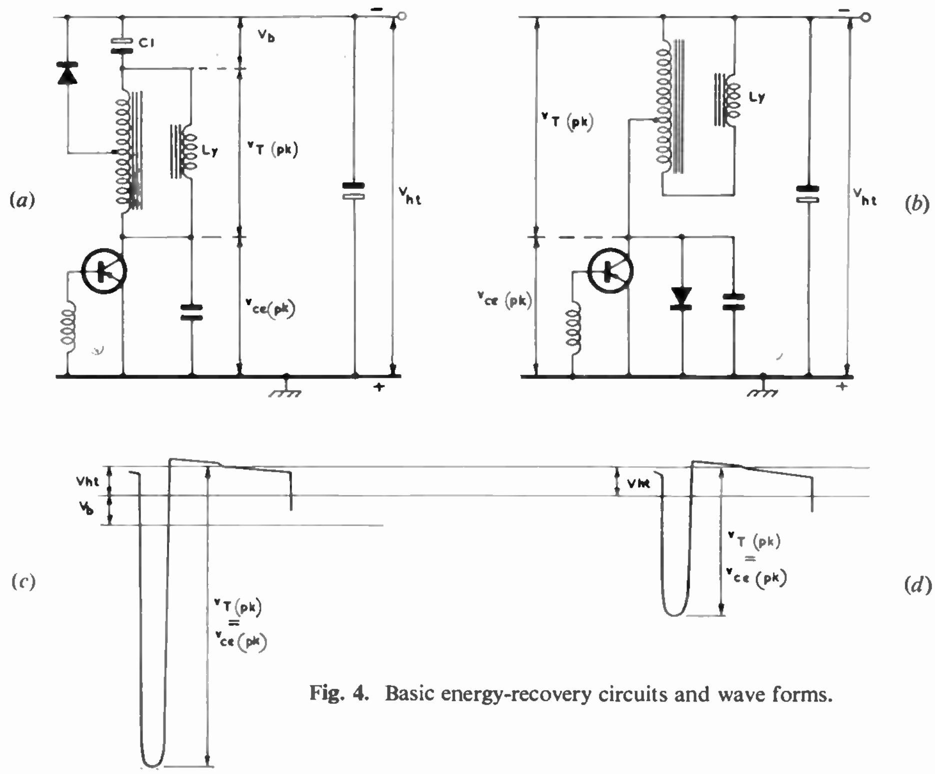


Fig. 4. Basic energy-recovery circuits and wave forms.

of the total scanning period and the h.t. line was 12·6 V then, from eqn. (7), $v_{ce(pk)}$ would be too large for present-day transistors, taking into account production spreads of components and circuit operational variations. Thus, the shortest flyback ratio for a given h.t. potential is determined by the peak collector-to-emitter voltage rating of the output transistor.

Determining the longest flyback period is not so definite. Modern picture tubes have a phosphor-coated area giving a picture with approximately a 4 : 5 aspect ratio. As the transmitted picture has a 3 : 4 aspect ratio, if the field height is adjusted to cover the phosphor height, about 6% over-scan must be allowed on line width to give a linear picture. This means that the line flyback period can be lengthened without any visible loss in scan until a flyback period of 21·4% is reached. Then the phosphor is just filled by the raster. If the flyback time is made still longer, then the aspect ratio will be greater than 4 : 5, and for a given line scan, the field amplitude will have to be increased to give a linear picture. Also, as the flyback time is increased, more and more picture information is lost. At long flyback times, it can be difficult to generate the required e.h.t. with correct tuning. In general, the shortest flyback period once the transmitted blanking period has been exceeded, consistent with $v_{ce(pk)}$ is the most desirable, and 22% is a recommended figure, although for some designs it may be found desirable to increase this slightly.

6. Circuit Design

The design problems are very similar to those of a valve line time-base. Although the efficiency of the valve circuit is an important factor, it is quite often sacrificed in the interests of economy. While in the transistor circuit cost is still an important consideration, efficiency is of major importance. Since the line time-base consumes the major part of the power of all circuits, it will determine to a large extent the viewing time of the receiver when operated from batteries.

6.1. *Scanning Coil Inductance*

It is desirable for maximum efficiency to choose the scanning coil impedance such that the coils are "directly driven", in fact, connected directly between the h.t. line and collector. For a 90 deg. deflection circuit, L_y will be about 140 μH. Even with the two coils connected in parallel, these values would mean that the coils will need to be wound with much thicker wire than has hitherto been used. But thicker wire is more difficult to wind and results in poor packing. Thus, to avoid the thicker wire, coils have been wound with several strands of thinner wire used together.

To obtain 140 μH coils for the 90 deg. application four wires would have to be wound in parallel. It is possible that some manufacturers would prefer to keep the number of strands to a minimum for ease of winding, yet on the other hand the greater the number of strands the fewer turns on the coil and the shorter the winding time.

For the design described in this paper only two wires have been wound in parallel, which gives a coil inductance of 550 μH. This only requires a 2-to-1 turns ratio between the secondary (to which the scanning coils are connected) and the primary winding (h.t. to collector). With this order of turns ratio, it is still possible to obtain tight coupling between the primary and secondary windings.

6.2. *Output Transformer*

The output transformer is a component in which excessive scanning power can be lost unless considerable care is taken in its design. Calculations of turns ratio will be given in later sections of this paper, and this section is mainly concerned with the layout and methods of winding.

6.2.1. Secondary inductance and coupling factor

It has been shown by Friend[1] that there is an optimum ratio of secondary to scanning-coil inductance for a given coupling factor k. As k increases, the optimum value of L_s increases, as does the efficiency of the transformer. Friend defines the ratio of the values of peak-to-peak ampere-turns of the yoke (*a*) when fed by a practical transformer and (*b*) when fed by an ideal transformer, as being a deflection factor F_D, given by

$$F_D = \frac{1}{\left[\frac{L_s}{L_y}\left\{\frac{1}{k^2}-1\right\}+\left\{\frac{2}{k^2}-1\right\}+\frac{1}{k^2(L_s/L_y)}\right]^{\frac{1}{2}}} \quad \ldots\ldots(8)$$

Figure 6 shows a transformer with a primary winding of 22 turns and a secondary winding of 86 turns with various arrangements of connections and winding configurations. The coupling between the windings in Fig. 6 (*a*) is 0·972, but this only gives a deflection factor of 0·735, which means a loss of 26·5% in scan. Figures 6 (*e*) and 6 (*f*) give the highest value of F_D, that is 0·933, which corresponds to a loss of scan of only 6·7%. Therefore, by choosing two methods of winding the same number of turns, the scan amplitude can vary by some 20%. Figures 6 (*d*), 6 (*e*) and 6 (*f*) show the best method of winding, where all layers of the winding are about the same width as the primary winding. It is not very important which layer is used for the primary. This can be determined by other circuit requirements, which are discussed in the following section.

6.2.2. Arrangement of connections

If, for a particular application, the scanning-coil inductance can be made such that the number of

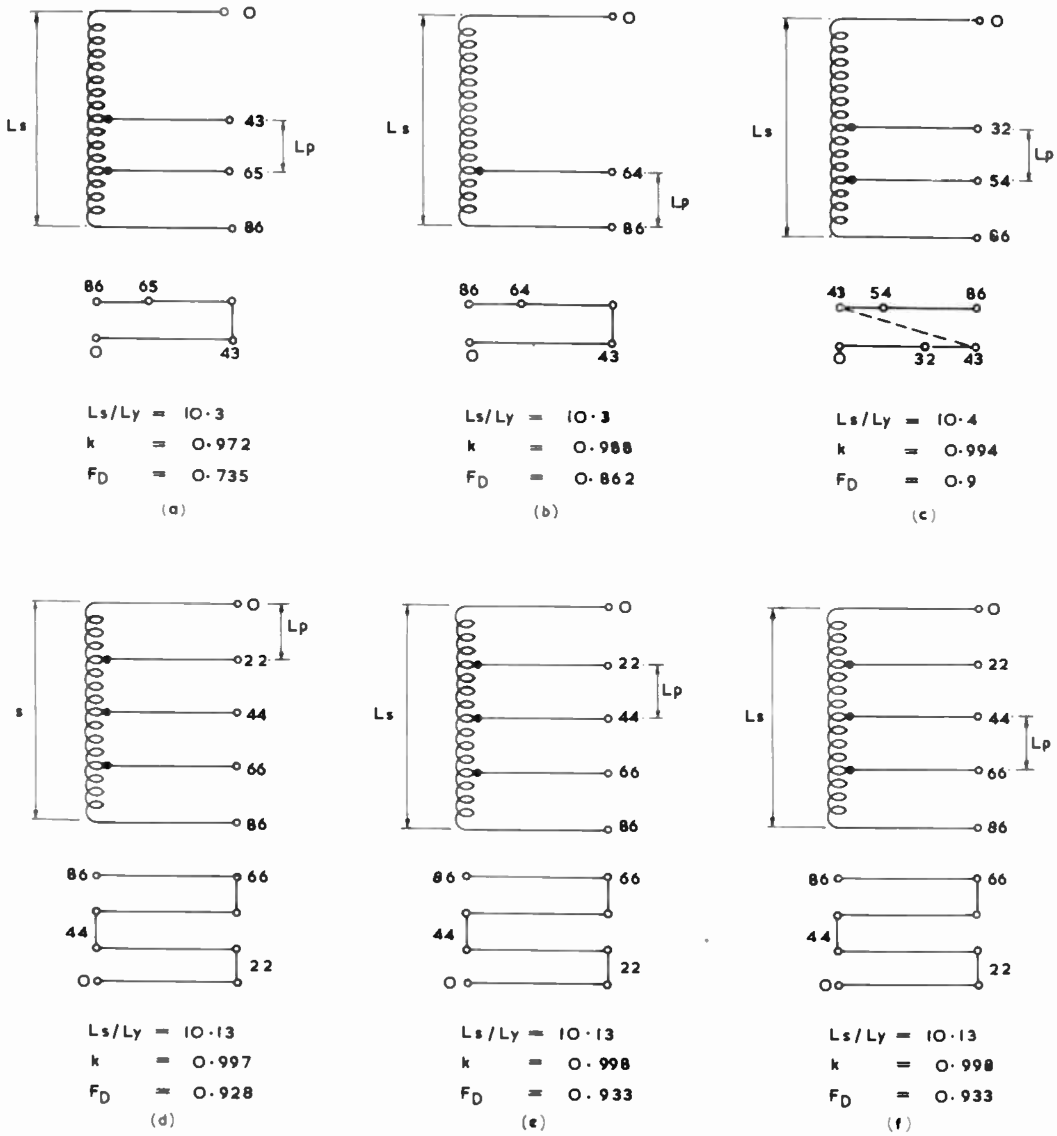


Fig. 6. Transformer winding arrangements.

secondary turns is twice the number of primary, then, depending on the ease of winding, the circuit can be connected as shown in Fig. 7. Point A will have a positive-going pulse voltage with respect to earth and point B a negative-going pulse voltage. These pulses are useful for the flywheel discriminator circuit after being fed through suitable capacitors.

The voltage pulse at point B may be rectified to give the video amplifier h.t. supply. Positive potentials will be required for first anode and focus supplies which are between 300 and 400 V. As will be seen from Fig. 7, the positive excursion of the voltage waveform $v_{pk(pos)}$ at the voltage at point A, is less than the negative excursion $v_{(pk)}$ of the voltage at the point B, the relationship being

$$v_{pk(pos)} = v_{pk} - nV_{ht} \qquad \ldots\ldots(9)$$

where n is n_{sec}/n_{pri}. Also

$$v_{pk} = FV_{ht}(n-1) \qquad \ldots\ldots(10)$$

From eqns. (9) and (10)

$$FV_{ht}(n-1) = v_{pk(pos)} + nV_{ht}$$

that is, $$v_{pk(pos)} = FV_{ht}(n-1) - nV_{ht}$$

whence, $$n = \frac{v_{pk(pos)} + FV_{ht}}{V_{ht}(F-1)} \qquad \ldots\ldots(11)$$

Thus for a required value of $v_{pk(pos)}$ (to be rectified for the first anode and focus supply), the required number of turns may be calculated. It is more convenient to obtain this positive voltage with a winding on the primary side of the transformer than by tapping the e.h.t. winding, since this is usually wave-wound.

6.3. *Non-Linearity*

The non-linearity of the time-base will depend on the voltages across the resistive elements in the circuit. These are the saturation resistance of the output transistor, the forward resistance of the efficiency diode, the resistance of the scanning coils, the impedance of the power supply and the resistance of leads and wiring. If the voltages are referred to the primary when the scanning coils are connected between h.t. supply and collector, and the efficiency diode is connected between collector and earth then,

$$\text{non-linearity} = \frac{(V_{tr} + V_{df} + V_{yr} + V_{sr} + V_w)}{V_{ht}} \times 100\%$$

For a practical design, the voltage drop V_{tr} across the transistor is about 0·5 V and the voltage V_{df} across the diode 0·5 V. Because of the impedance at line frequency of the large electrolytic capacitors decoupling the h.t. line, a sawtooth voltage appears across the capacitor. The polarity of this voltage is such that the voltage across the transformer is higher than the h.t. line voltage at the start of scan when current is fed into the supply by the efficiency diode, and lower during the second half of scan when the transistor is drawing current from the supply. A sawtooth voltage amplitude of 0·25 V has been measured in a typical circuit. It is difficult to measure the resistance of the circuit wiring, but it is very important to keep this as low as possible. The non-linearity measured in a typical circuit is 25%. If V_w is ignored, the calculated value is equal to

$$100(0{\cdot}5+0{\cdot}5+1{\cdot}4+0{\cdot}25)/12{\cdot}6$$

or approximately 21%. The difference between the measured and calculated values is therefore caused by V_w.

Non-linearity correction can be obtained by conventional means, either with the "saturated reactor" or the "short-circuited turn". With valve circuits the latter method has become popular mainly because of its low cost and simplicity. The saturated reactor tended to give rise to "ringing" on the scan. The natural frequency of the oscillations, determined by its inductance and the circuit stray capacitance, was quite low and an oscillatory voltage appeared across the scanning coils. The inductance of the line coils used for transistor time-bases is much lower giving a lower inductance for the saturated reactor. The stray capacitance of the circuit is also much lower and the resultant oscillations are at a very much higher frequency. The major part of the circuit stray capacitance is across the scanning coils; therefore very little oscillatory voltage appears across them, and it has

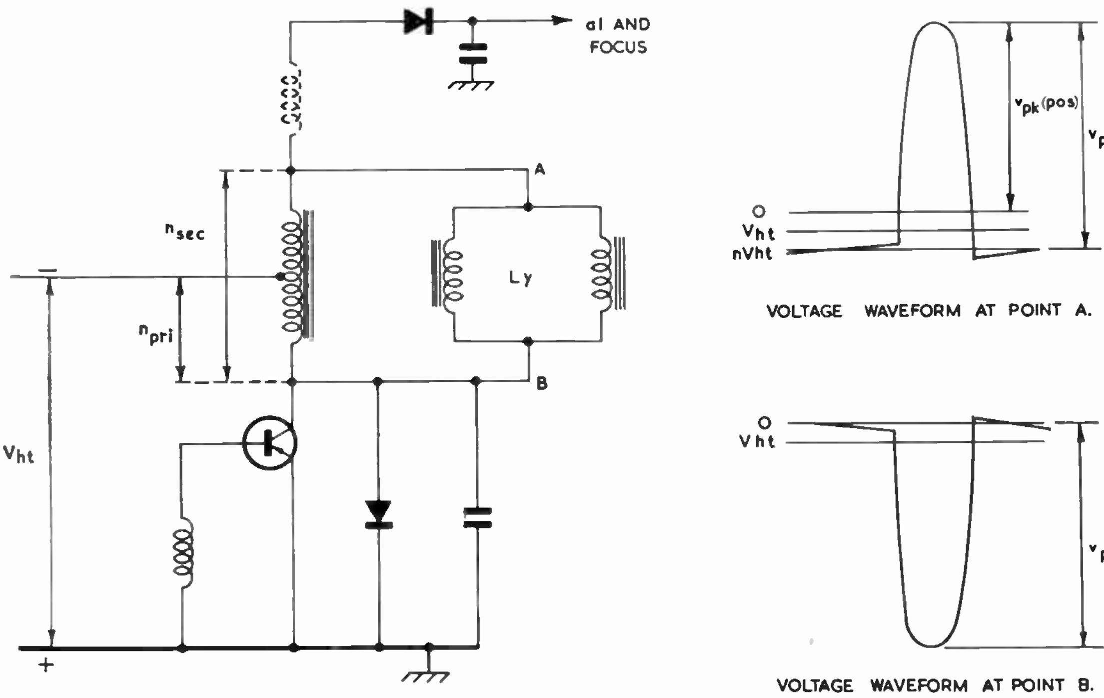


Fig. 7. Balanced scanning coil connections.

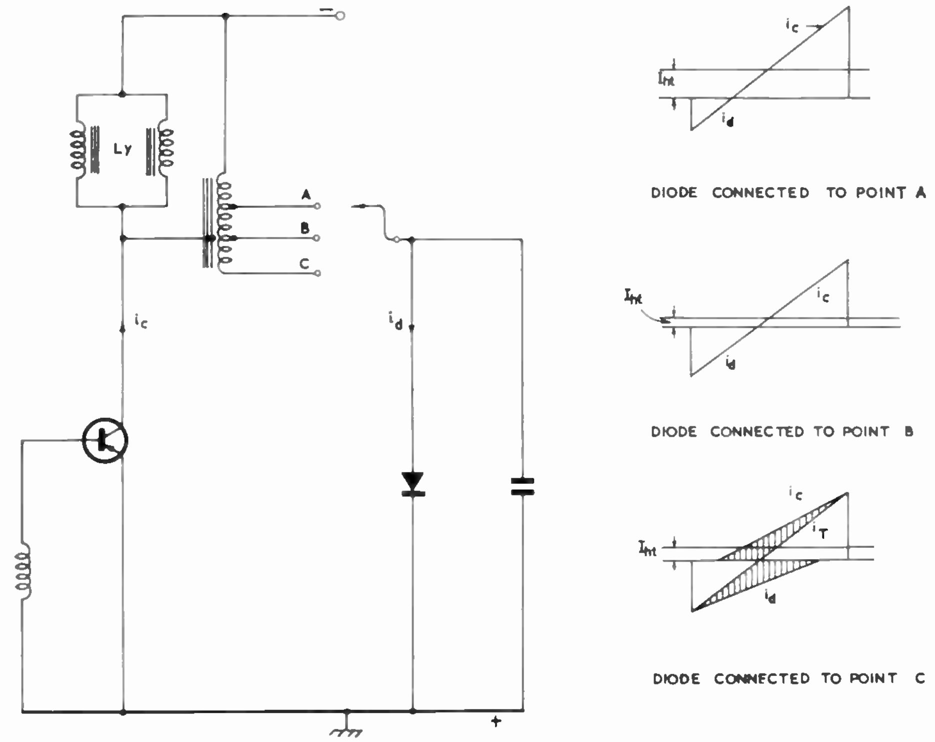


Fig. 8.
Effect of diode connections.

been found unnecessary to place any damping components across the saturated reactor.

It has been found that using the saturated reactor the circuit efficiency is higher than with the short-circuited-turn method. There is no power lost due to damping components, and the energy stored in the device at the end of the scanning stroke can be recovered.

6.4. *Efficiency Diode Connection*

The effect of varying the point where the diode is connected to the transformer is shown in Fig. 8. Connection of the diode to point A, that is, across fewer turns, leads to a diode conduction period which is short compared with the transistor conduction period. This gives high peak collector currents with, consequently, a large h.t. current.

Connection of the diode to point B, that is, the same point as the collector, gives the most efficient arrangement. Then the diode and transistor conduction periods are nearly equal, equality of the conduction periods giving the highest circuit efficiency.

With the diode connected to point C, both the diode and transistor conduction periods are lengthened, which leads to increased circulating energy in the circuit not contributing to the scan. This additional energy (shown as the shaded portion in the diagram) gives rise to increased losses and a lower efficiency. The non-linearity of scan will, however, be reduced because the diode impedance referred to the primary winding is lower. Although less linearity correction is required (with, therefore, a reduction in the loss of scan), the overall efficiency is still lower than when the diode is connected to the collector. If the efficiency diode has an appreciably higher forward resistance than the saturation resistance of the transistor, then this type of connection can be used to some advantage. Even though the circuit is inherently less efficient, the required linearity correction is less and the overall performance may be improved.

6.5. *Width Adjustment*

A width control which does not affect the e.h.t. potential, flyback time and circuit efficiency is an ideal which cannot be achieved in practice. Series or parallel inductors in the scanning-coil circuit, although simple arrangements, have all the disadvantages given above.

Due to the spread of components in production, the width can change by $\pm 5\%$. Having too much overscan is not as undesirable as having too little and it is really only important to cater for the -5% condition. It does not seem worth while having a width control to cover this small range especially considering the difficulties in designing width controls which do not detune the transformer. Therefore, in the 90 deg. receiver design no width control has been incorporated, but the scan at the nominal h.t. potential of 12·6 V has been designed to be 15% greater than the horizontal phosphor dimensions of the picture tube. With limit components giving the low scan conditions, there will still be 10% overscan, which will allow the h.t. potential to fall by some 20% during the battery discharge before the edge of the raster becomes visible.

6.6. *Third Harmonic Tuning*

In Fig. 5 it was shown that for a 22% flyback period and using 3rd harmonic tuning, $v_{ce(pk)}$ was equal to 77 V. If 3rd harmonic tuning is not used in the transformer, then $v_{ce(pk)}$ will be greater for a given h.t. potential and flyback period, and either the flyback period must be increased or the h.t. potential reduced.

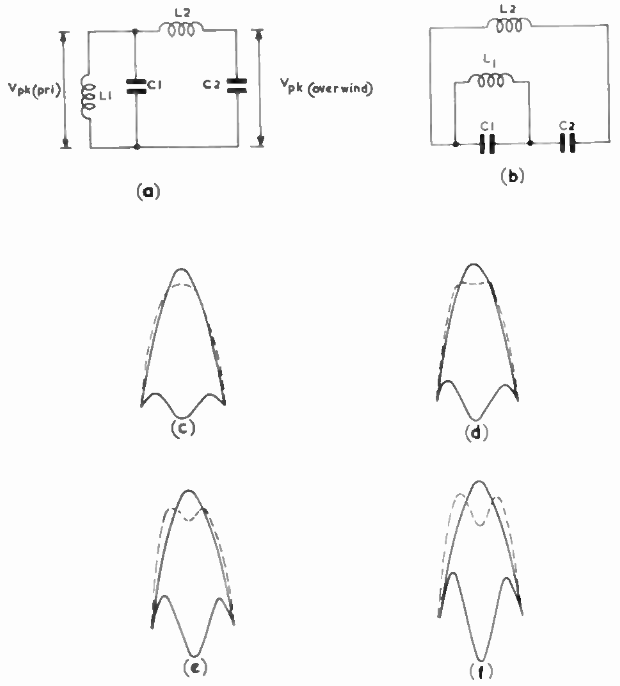


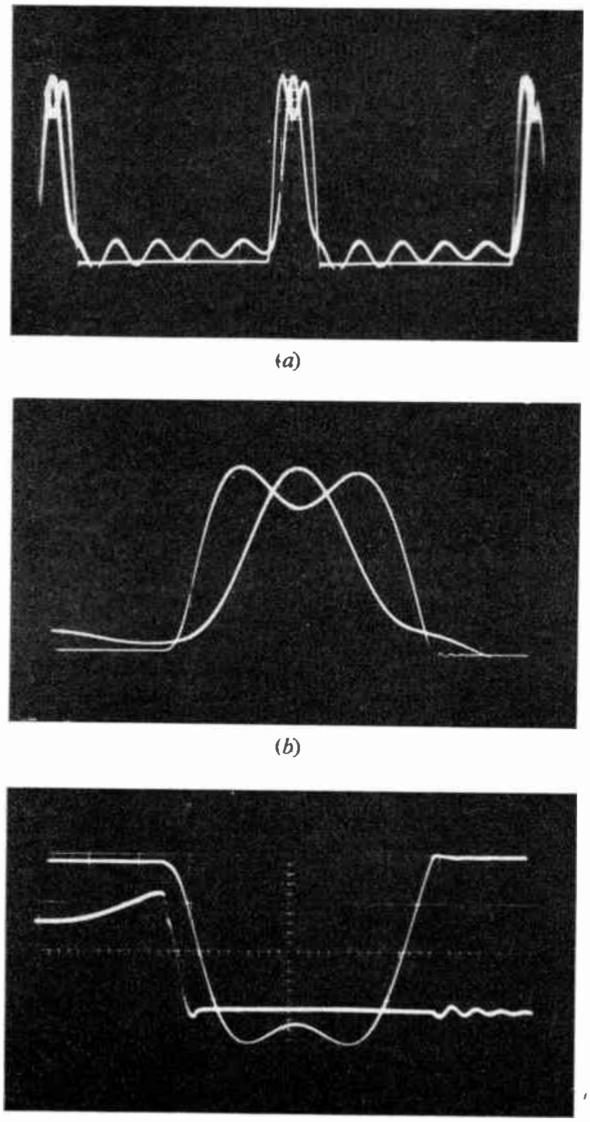
Fig. 9. Equivalent circuit and peak primary voltages during flyback.

It is now common practice with valve receivers to tune line transformers. It has been reasonably easy to obtain the correct values of leakage inductance between the primary and e.h.t. windings and also the value of self-capacitance of the e.h.t. winding such that the tuning is correct. The number of turns in the e.h.t. overwind and the primary winding are nearly equal and, in general, potentials of 6 kV and 10 kV appear across the primary and the overwind respectively.

With transistor circuits, the position is very different. A potential of only about 77 V appears across the primary winding and, furthermore, it is a negative pulse, so that no contribution towards the e.h.t. potential can be obtained from the primary. This means that the full e.h.t. potential must be generated by the overwind. Whereas with valve circuits the primary winding might consist of 1100 turns and the overwind of 1300 turns, with transistor time-bases the overwind will be made up of the sum of these turns. This will give a large winding with an appreciable self-capacitance. Unless great care is taken, the "ringing" frequency can be too low for the desired flyback frequency, even though the flyback period has been lengthened to 22% of the total line period. The overwind has to be wave-wound to give sufficiently low self-capacitance.

Figure 9 (*a*) shows the equivalent circuit of the transformer during the flyback period. L1 is the inductance of the transformer and scanning coils and C1 is the associated capacitance. L2 is the leakage inductance between the primary and the overwind, and C2 is that self-capacitance of the overwind which is not reflected into the primary winding ($k < 1$). L1 and C1 are the main components giving the flyback time, and L2 and C2 mainly determine the 3rd harmonic frequency (ringing frequency). If Fig. 9 (*a*) is redrawn as in Fig. 9 (*b*), it will be seen that ringing voltages developed across L2 will appear across both C1 and C2, the amplitudes depending on the value of the capacitances. If the value of C2 is too small compared with C1, insufficient peak voltage reduction will be obtained (Fig. 9 (*c*)). Increasing the value of C2 (which implies a smaller value of L2 to maintain the correct frequency of the ringing) will increase the ringing voltage developed across the primary (Fig. 9 (*d*)). This condition gives the maximum reduction of 15% in $v_{ce(pk)}$. Increasing C2 still further gives the condition shown in Fig. 9 (*e*), where again the maximum reduction of $v_{ce(pk)}$ is achieved. Figure 9 (*f*) shows the condition where C2 is too large with respect to C1. The ringing voltage across the primary is then too great, and maximum reduction in $v_{ce(pk)}$ is not obtained. The most satisfactory conditions of operation are shown in Figs. 9 (*d*) and 9 (*e*), where the maximum reduction in the primary voltage is achieved. However, the condition shown in Fig. 9 (*e*) is the optimum since a higher peak voltage is developed across the overwind because of the larger voltage in the circuit, and fewer turns are required for a given e.h.t. potential. There is, therefore, an optimum ratio of C1 to C2 to give the most satisfactory circuit conditions. It has been found in practice that the overwind can have up to 60% greater peak voltage per turn than the primary when tuned to this optimum condition.

Optimum tuning is best found in practice by trial and error. One method is to wind two complete primary windings, one on each round limb of the transformer core, with the overwind placed on top of one of these windings (called the "coupling" winding). The other winding is connected into the



(a)

(b)

(c)

Fig. 10. Collector and overwind voltages during flyback period.

circuit as a normal primary. One end of the primary is connected to one end of the coupling winding, and the other ends of the two windings are connected together with a small variable inductance. This gives two identical windings connected in parallel through the variable inductance. This inductance has an effect similar to that of a leakage between the primary and the overwind.

The flyback period is adjusted to its required value by means of a capacitor across the primary winding. The variable inductance is changed until the correct tuning condition is obtained. This can be recognized by use of a double-beam oscilloscope where on one beam the primary voltage is displayed and on the other beam the voltage waveform across the overwind is displayed. This can be obtained by placing the oscilloscope probe close to the winding which will pick up sufficient voltage to be displayed. The tuning is adjusted until the primary voltage waveform is as in Fig. 9 (*e*) with symmetrical "humps" and the maximum voltage reduction is obtained. It has been found that this is not necessarily the minimum ringing condition, which is given when the first hump is larger than the second, but is quite close to it. These uneven humps are caused by the ringing voltage decaying during the flyback period. With symmetrical humps, a voltage reduction of about 15% is possible. Figures 10 (*a*) and 10 (*b*) are photographs of the waveforms of the primary and overwind voltages for the 90 deg. receiver design. From Fig. 10 (*a*) it will be seen that some ringing exists on the e.h.t. voltage waveform. Figure 10 (*b*) is an expanded view of Fig. 10 (*a*) and shows the 3rd harmonic content in the voltage waveforms.

If, when obtaining a tuned condition, the primary voltage waveform has not decreased by 15% from the untuned value and the waveform appears as in Fig. 9 (*c*), then a higher-capacitance overwind is required. If the reduction in primary voltage is insufficient and the waveform appears as in Fig. 9 (*f*), then a lower-capacitance overwind is needed. When the conditions as shown in Fig. 9 (*e*) are found, with the required reduction in primary peak voltage and the desired e.h.t. and flyback time, then the overwind capacitance is correct. It should not be forgotten that the overwind requires waxing, which will increase its capacitance, and it is best to wax the winding before making adjustments.

The correct leakage inductance must be obtained without the variable inductance. The primary winding can be tapped at about 3 or 5 turns and a winding with this number of turns placed on the limb containing the overwind. The tapped part of the primary winding is connected in parallel with the coupling winding. This winding is then moved along the limb, thus varying the coupling with the overwind, until the correct tuning is obtained. More than 3 or 5 turns may be required to obtain the tuned condition. The coupling winding is the same number of turns as are tapped on the primary.

It is important that the length of the finish lead of the overwind, which is connected to the anode of the e.h.t. rectifier, should be the same during all the adjustments. The capacitance to earth of this lead, which appears across the overwind, is very critical when tuning the transformers. A typical turns ratio between the primary and the overwind is 100 and the reflected capacitance of this lead appearing across the primary winding is therefore n^2kC, or $kC.10^4$. The capacitance to earth of this lead need only vary by a few picofarads to make a difference to the primary capacitance of several hundredths of a microfarad. The lead capacitance will also affect C2 in Figs. 9 (*a*) and 9 (*b*), and unless consistency in length is maintained, it can cause confusion when making adjustments.

6.7. *Line Transformer Cores*

Because of the large step-up between the primary and the overwind, it is desirable to keep the number of turns on the primary to a minimum. If a high ratio of L_s to L_y is to be maintained, then it is an advantage to use Ferroxcube U-cores with as large a cross-sectional area as is practical. With valve time-bases, the trend has been to reduce the cross-sectional area of the core, and to use a greater number of turns on the windings. Improved ferrite materials with a high saturation figure have made this possible. With transistor circuits it is advantageous to use less copper and more ferrite to maintain the circuit efficiency.

As the heat in a transistor receiver is very much less than in a valve receiver, the Ferroxcube will operate at a lower temperature. The peak flux density may therefore safely be higher, the working point being determined from the published data. This can mean in some designs a reduction in the air-gap, or possibly no gap at all, when low scanning power is required. This reduces still further the number of turns for a given inductance and thus reduces the size of the overwind.

7. Transistor Operating Conditions

Overloading semi-conductors and exceeding the ratings can cause catastrophic failure of the devices. With valves it has been usual to use the design centre rating system, but with semi-conductors the absolute maximum system is used. This rating means that under the worst probable operating conditions of the device, the ratings must not be exceeded.

The semi-conductors used in a line output stage are required to switch very large powers and unless care is taken in the design, the devices can be easily overloaded with the consequent risk of failure.

7.1. *Peak Collector Voltage*

It is extremely important that the rating for the peak collector voltage of the transistor and efficiency diode should not be exceeded under the worst probable conditions of circuit operation. When designing a circuit it is often desirable to meter this value continuously. A simple peak reading voltmeter is required, providing it does not load the circuit.

7.2. *Working Point*

During the scanning stroke, the transistor is operating below the knee of the I_c/V_c characteristic. This is the most efficient method of operation because it makes full use of the already low h.t. line potential, and reduces the collector dissipation to a minimum. The bottoming potential for a typical transistor at a peak collector current of 5 A is 0·5 V. When operating from a 12·6 V h.t. supply, the voltage across the transformer primary at the end of the scanning stroke is

$$\frac{100(12{\cdot}6-0{\cdot}5)\%}{12{\cdot}6} = 96\% \text{ of the h.t. potential}$$

A typical value for a valve circuit is

$$\frac{100(200-25)\%}{200} = 87{\cdot}5\% \text{ of the h.t. potential}$$

7.3. *Transistor Turn-off*

It is well known that transistors when passing large collector currents cannot switch from a saturated to open-circuit condition quickly.[2] During the scanning stroke, the output transistor is driven hard into saturation, and an excess carrier density builds up in the base. This must be removed before the collector current can reverse, and even though the base current reverses, the collector current will still increase in a saw-tooth manner. The collector current will not start to reverse until the carrier density in the base has decayed to a value just insufficient to bottom the transistor. The time taken from the reversal of the

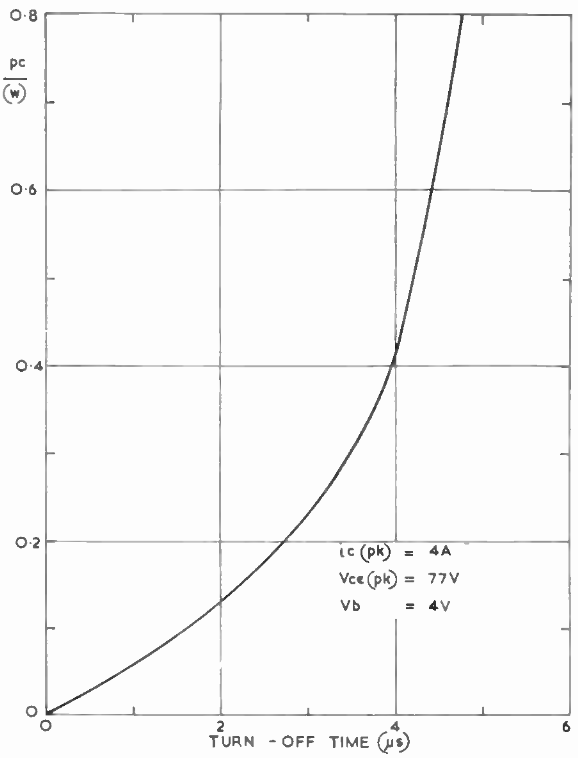


Fig. 11. Variation of collector dissipation with transistor turn-off time for circuit using 3rd harmonic tuning.

base current to the start of reversal of the collector current is called the storage time, and the time taken from the end of the storage time (when the flyback is initiated) to when the collector current is nearly zero is known as the decay time. Because of the storage time, direct synchronizing of the time-base cannot be used and flywheel synchronizing has to be employed.

During the decay time when the collector current is still flowing in the flyback period, there will be a power loss. If this time is excessive, large collector dissipation and a significant worsening of the circuit efficiency can occur, since power lost during the flyback period means less energy to be recovered during the first part of the scanning period. Figure 10 (*c*) shows the collector current and collector voltage during the flyback period and immediately after the start of the flyback period when there is still a large collector current flowing. The voltage amplitude is reduced compared with when no current is flowing. This can be seen by comparing the slopes of the voltage waveform at the start and finish of the flyback period.

In Fig. 11, collector dissipation is plotted against decay time. It will be seen that up to a decay time of about 3·5 μs, the collector dissipation is small, but then increases rapidly. Obviously the longer the flyback period the lower will be the dissipation.

7.4. *Junction Temperature*

Absolute maximum ratings of junction temperature for transistors used in line output stages is of the order of 90°C. This must not be exceeded when components and transistors are at their limit values. The power dissipated in the transistor arises from,

(*a*) decay time

(*b*) resistance of collector-emitter junction

(*c*) leakage current

(*d*) resistance of base-emitter junction.

Usually the main loss in the transistor is due to the decay time but, of course, this will depend on the transistor and the flyback time. Obviously, the size and nature of the heat sink will also determine the junction temperature. It is usual to choose the size of heat sink such that when all components are at their limiting values, during most unfavourable operational conditions (high h.t. line, maximum beam current and "out of synch") and the limit characteristics for the transistor and diode, the junction temperature will not exceed the published value. A larger heat sink than gives this value is obviously an advantage, but space limitation may prevent this.

7.5. *Avalanche Multiplication*

It could be possible in a design to use a very large heat sink and therefore, due to the lower junction temperature, employ a transistor having a longer decay time. It is important then to ensure that the operating point of the transistor during the decay time does not enter the region where avalanche multiplication can occur. If this occurs for only a few microseconds a very large instantaneous power will be dissipated in the transistor which could cause a failure.

7.6. *Base-Emitter Drive*

Drive circuits are discussed in an accompanying paper[3] and will not be described in detail. Briefly, it is important that the line output transistor must be bottomed during its conduction period and the drive must ensure that it remains in this state with limit components and limit operation conditions, otherwise large collector dissipations can occur.

During the flyback period the transistor must be turned off as quickly as possible and this is achieved by applying a large reverse bias to the base emitter diode.

8. Time-base Design for a 90 deg. Picture Tube (12 kV E.H.T.)

This section describes a line time-base scanning a 90 deg. deflection picture tube with a narrow neck at 12 kV. The circuit is designed to operate from a nominal h.t. potential of 12·6 V with a variation of ±20%. Allowance must be made for this large variation to cater for the normal charge/discharge characteristics of batteries.

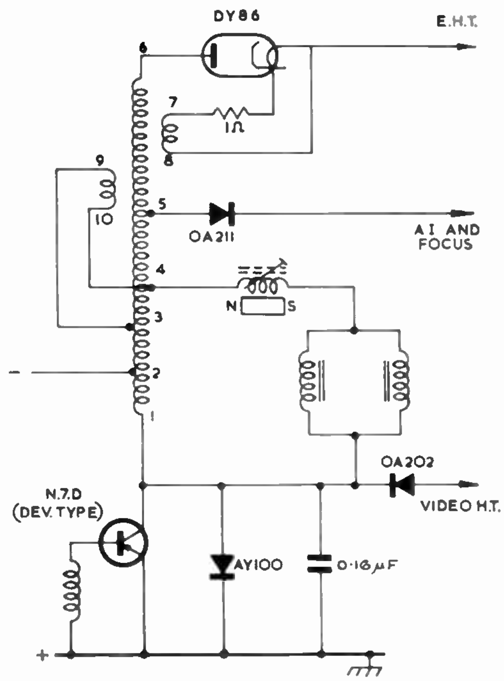


Fig. 12. Circuit diagram of line time-base for 90 deg. tube.

8.1. *Circuit Description*

Figure 12 shows the circuit diagram of the time-base. Winding 1–2 is the primary and 1–4 the secondary. To obtain the correct 3rd harmonic tuning, a coupling winding 9–10 is wound by the side of the overwind 5–6 and coupled into the secondary across taps 3 and 4. Winding 4–5 is to step up the fly-back pulse to give the tube first anode and focus supply. The h.t. supply for the video amplifier is taken from point 1. The capacitor connected across the efficiency diode (AY100) is the tuning capacitor to obtain the correct flyback time. It will be seen that this has a value of 0·16 μF, which is non-standard. This is in fact the correct value to give 12 kV e.h.t. potential with a 22% flyback period, and $v_{ce(pk)}$ is 77 V. If a capacitance value of 0·15 μF is chosen, the e.h.t. potential is 12·4 kV, the flyback period 21·6%, and $v_{ce(pk)}$ is 79 V, measurements have been made at the 22% flyback period which give an extra slight margin of safety on $v_{ce(pk)}$.

No "S-correction" is provided in the design to counteract the effect of the picture-tube face-plate curvature. There are two reasons for this, firstly it is difficult to design a satisfactory flywheel circuit with a capacitor in series with the scanning coils, and secondly, if a capacitor is used, it requires a value of 8 μF, which means it would have to be an electrolytic. The polarizing voltage would only be about 0·1 V, and furthermore, it is doubtful if the current handling ability of the capacitors would be adequate. Unless the resistance of the capacitor was very low, appreciable loss would result. Without S-correction the non-linearity measured from the edges to the centre of the raster is 10% and, as this is a symmetrical change, it is not objectionable.

8.2. *Circuit Design*

For a 22% flyback period, $V_{ht} = 12{\cdot}6$ V, and with 3rd harmonic tuning

$$v_{ce(pk)} = \frac{0{\cdot}85\pi t_t V_{ht}}{2t_f}$$

$$= \frac{0{\cdot}85 \times 100 \times 12{\cdot}6 \times \pi}{2 \times 22}$$

$$= 77\,\text{V}$$

If the inductance of the saturated reactor is taken as $L_y/10$, then

$$V_{sec} = \frac{1{\cdot}1\, L_y\, i_{y(pk-pk)}}{t_s}$$

$$= \frac{1{\cdot}1 \times 555 \times 10^{-6} \times 3{\cdot}36}{77 \times 10^{-6}}$$

$$= 26{\cdot}7\,\text{V}$$

The turns ratio n is given by

$$n = \frac{V_{sec}}{V_{pri}}$$

$$= \frac{26{\cdot}7}{12{\cdot}6}$$

$$= 2{\cdot}12$$

From Table 1, the transistor peak VA is 325, therefore

$$i_{c(pk)} = \frac{325}{77}$$

$$= 4{\cdot}22\,\text{A}$$

The peak diode current $i_{d(pk)}$ is given by

$$i_{d(pk)} = 0{\cdot}8\, i_{c(pk)}$$

$$= 0{\cdot}8 \times 4{\cdot}22$$

$$= 3{\cdot}38\,\text{A}$$

From Fig. 2,

$$I_{ht}\, t_t = \frac{i_{c(pk)}t_c}{2} - \frac{i_{d(pk)}t_d}{2} \qquad \text{......(12)}$$

where t_c and t_d are the conduction times of the transistor and diode respectively.

If,

$$\frac{i_{d(pk)}}{i_{c(pk)}} = \frac{t_d}{t_c} = \beta$$

Then from (12),

$$I_{ht} = \frac{i_{c(pk)}t_s(1-\beta)}{2t_t}$$

For $\beta = 0{\cdot}8$ and for a 22% flyback period,

$$I_{ht} = \frac{4{\cdot}22 \times 77 \times 10^{-6} \times 0{\cdot}2}{2 \times 100 \times 10^{-6}}\ \text{amps}$$

$$= 325\,\text{mA}$$

(This calculated value is for the circuit to provide the required current in the scanning coils only, and it will increase because of e.h.t. power, the heaters and the series resistance for e.h.t. rectifier and h.t. supplies to other circuits in the receiver.) Therefore the power to scan the picture tube is $0{\cdot}325 \times 12{\cdot}6 = 4{\cdot}1$ W. Additional power requirements will be as follows:

E.h.t. power at 100 μA beam current	= 1·17 W
Video, first anode and focus supplies	= 0·42 W
DY86 heaters plus 1 Ω series resistor	= 1·07 W

The total additional power is therefore 2·66 W. If an allowance is made for an 80% conversion efficiency by the time-base, then the additional power requirement is 3·32 W. Therefore the total power for time-base at 100 μA beam current is (3·32+4·1) or 7·42 W. It is interesting to note that only 55% of the total power taken by the time-base is required to provide the scan.

8.3. *Circuit Measurements*

The results of circuit measurements are given in Table 2. When the h.t. line voltage is 12·6 V and the picture tube beam current is 50 μA, then

Video h.t. supply	= 75 V at 5 mA
First anode and focus supply	= 400 V at 100 μA
Flyback period	= 22%

The scan on the AW36/10 is 15% greater than horizontal phosphor dimension. A suitable heat sink for both the N7D and the AY100 would be 4×4 in. of $\frac{1}{16}$ in. non-blackened aluminium.

Table 2

Circuit Measurements

V_{ht} (V)	I_b (μA)	E.H.T. (kV)	$v_{ce(pk)}$ (V)	$i_{c(pk)}$ (A)	I_{ht} (mA)	P_{ht} (W)
10·0	0	10·1	62	3·3	410	4·1
10·0	50	9·8	64	3·4	460	4·6
10·0	100	9·5	66	3·5	505	5·05
10·0	150	9·3	67	3·55	540	5·4
12·6	0	12·2	77	4·2	505	6·36
12·6	50	11·9	80	4·3	560	7·06
12·6	100	11·7	83	4·4	600	7·56
12·6	150	11·5	85	4·45	640	8·06
15·0	0	14·6	95	5·0	600	9·0
15·0	50	14·3	98	5·15	655	9·82
15·0	100	14·0	100	5·3	705	10·58
15·0	150	13·7	102	5·5	750	11·25

8.4. *Output Transformer and Linearity Control*

8.4.1. Output transformer winding specification

Winding 4–1 : 56 turns tapped at 5 and 30 turns of 27 s.w.g. Lewmex F wire; layer-wound in two layers.

Winding 4–5: 145 turns of 38 s.w.g. Lewmex F wire; layer-wound in one layer.

Winding 5–6: 2700 turns of 42 s.w.g. enamelled copper s.s.c. wire, wave-wound, 5 mm wide.

Gears on Douglas wave-winder:

A	B	C	D	E	F
39	34	32	36	60	60

Winding 7–8: 2 turns of suitably insulated wire.

Winding 9–10: 5 turns of 27 s.w.g. Lewmex F.

Winding 4–5 is wound on the former first, followed by winding 4–1. Winding 5–6 is placed on the opposite limb of the core to windings 1–5. Winding 9–10 is placed on the same limb as winding 5–6, the two windings being spaced $\frac{9}{16}$ in. apart. Winding 7–8 is wound on one of the side limbs of the cores. It is important to arrange the direction of windings 1–4, 4–5, 5–6 and 9–10 such that the voltages are in the correct sense.

Cores: Two FX2351 assembled with total gap of 0·002 in.

Formers for windings 1–5 and 5–6 plus 9–10: length 1·25 in., internal diameter 0·67 in., wall thickness $\frac{1}{32}$ in.

8.4.2. Linearity control (winding specification)

100 turns, 24 s.w.g. Lewmex F wire, layer-wound in two layers.

Core: FX1054.

Former: length 2 in., internal diameter $\frac{3}{16}$ in., wall thickness $\frac{1}{16}$ in.

Magnet: M2529.

The magnet is spaced $\frac{1}{16}$ in. from the windings.

9. Acknowledgments

The author would like to acknowledge the contribution made by Mr. J. B. Hughes to the work described in this paper.

10. References

1. A. W. Friend, "Television deflection circuits: Part 2. Theory and design of combined low-loss horizontal deflecting and high-voltage power supply systems", *R.C.A. Review*, 8, No. 1, pp. 115-38, March 1947.
2. J. L. Moll, "Large-signal transient response of junction transistors", *Proc. Inst. Radio Engrs*, 42, pp. 1773-84, December 1954.
3. M. C. Gander and P. L. Mothersole, "Transistor video amplifier and line time-base synchronization circuits for television receivers", *J. Brit.I.R.E.*, 23, pp. 445–58, June 1962.

Manuscript received by the Institution on 10th March 1962 (Paper No. 736/T12).

U.D.C. 621.376.3/4: 621.394.3: 681.39

A Comparison of the Merits of Phase and Frequency Modulation for Medium Speed Serial Binary Digital Data Transmission over Telephone Lines

By

F. G. JENKS, B.Sc.†

AND

D. C. HANNON, B.Sc.†

Presented at the Symposium on "Data Transmission" in London on 3rd January 1962.

Summary: The telephone line is by no means an ideal channel for the transmission of serial binary digital data. It is important therefore to choose the method of transmission which is least affected by the various signal impairments which may be encountered. This paper indicates briefly why phase and frequency modulation are the best alternatives for general use and then examines in detail the relative merits of these two methods of modulation. The conclusions drawn from this study have been substantiated by measurements of error rate carried out on a number of p.m. and f.m. receivers working under conditions where comparisons could validly be made; the more important results are quoted in this paper. It is clearly shown that phase modulation has significant advantages over frequency modulation in the data transmission application.

1. Introduction

Because speech has low information content, a high degree of redundancy and tolerance to relative phase shifts, audio signals do not make exacting demands on the properties of the telephone channel. This fact has been exploited by telephone engineers in the interests of economy and as a result features are present which add to the difficulties of data transmission. A detailed consideration of these properties in relation to the proposed signals is given in succeeding Sections but, to provide a background to the problem, the important features of the channel are first outlined.

To keep cross-talk between channels to an acceptable level signals must be limited in respect of peak power, the normal transmitter peak power level being about −7 dBm. The signal is attenuated in transmission to an extent depending on the type of link, the length of the connection and the number of repeaters on the link. A receiver should be able at least to cope with line attenuations over a range from 0 to 30 dB.

The telephone channel has a frequency response which covers only the more significant audio frequencies. Whilst it is commonly assumed that the available bandwidth is from 300 c/s to 3000 c/s, lines having appreciably less bandwidth than this are frequently encountered. The delay characteristic of the line, however, usually imposes more severe limitations on data transmission than the amplitude response. The group delay curve is roughly parabolic in shape. Measurements on the British network by Post Office engineers show that on audio amplified lines, 100 miles in length, the 1 ms delay bandwidth (i.e. the bandwidth between the two frequencies having delays relative to the minimum of 1 ms) may vary from 1070 to 2100 c/s. On carrier links the worst 1 ms delay bandwidth reported by the same workers is 2300 c/s. Alexander, Gryb and Nast[1] report that in the United States of America the 1 ms delay bandwidth on 10% of long haul circuits is less than 1700 c/s.

Except under fault conditions white noise is rarely objectionable on telephone lines. On the other hand impulsive noise, caused by excitations of comparatively short duration which can be traced to lightning, ignition interference, dialling and switching transients and similar effects, presents serious problems. At the receiver such impulses are frequently oscillatory in character. They may last up to some tens of milliseconds and may exceed the signal in amplitude. They tend to occur in bursts and so produce short periods of very high error rate. Such noise is difficult to specify mathematically and an attempt to obtain statistical information experimentally gives consistent results only if the period of testing covers several months.[2] Cross-talk effects also arise in which the single tones used for control purposes and other signals are picked up due to coupling between adjacent circuits.

A further potential source of difficulty arises when carrier links are encountered. Here single-sideband transmission is used and because asynchronous modulating and demodulating carriers are usually employed, the signal may be translated in frequency by a few cycles per second. The modulating and

† The Plessey Company Ltd., Roke Manor, near Romsey, Hampshire.

demodulating carriers are usually referred to a frequency standard and corrections are applied to them if the error exceeds a prescribed limit. The rate at which such corrections are applied is also important since one type of receiver (see Sections 2.2 and 2.5) is sensitive to this effect.

Because the telephone channel has a low frequency cut-off, data cannot be satisfactorily transmitted in the form of a base-band signal using two d.c. states. Voltage transitions corresponding to changes of state of such a signal could be detected but this arrangement would be particularly susceptible to impulsive noise and the considerable phase distortion associated with the low frequency cut-off would also present a substantial hazard.

The data must therefore be used to modulate a carrier. Amplitude modulation has been used by some workers in this field.[3] For maximum efficiency the carrier needs to be 100% modulated and problems arise in applying a.g.c. to a signal which is absent for a large, variable proportion of the time. In addition because of cross-talk effects a telephone signal is limited in respect of maximum power and so an a.m. signal has a long-term average power level equal to only half the permitted maximum. The transmitted signal power is therefore 3 dB worse than for p.m. or f.m. where the maximum power level may be used continuously. Just as in the theory for 100% amplitude modulation by a sine wave, half the power in an a.m. binary-symmetric message signal appears as a carrier component which carries no signal intelligence. Thus the information-carrying sidebands have a power level 6 dB below that permissible on the channel. It will be shown later that p.m. has no wasted power of this kind but that f.m. does have about half its power in wasteful components which have a marked resemblance to the a.m. carrier.

Single-sideband amplitude modulation meets some of these objections and is potentially capable of high efficiency. However existing practical systems[4] on these lines do not by any means take full advantage of this possibility, and, indeed, the common practice appears to be to use in the receiver what is essentially an ordinary double-sideband a.m. detector. To make this rather curious arrangement work a large proportion of carrier has to be added to the sideband at the transmitter in order, as is usually said, to "reduce the quadrature component". In the authors' view this is a mistaken approach since it regards the s.s.b. signal merely as a distorted d.s.b. signal. Oxford[5] has recently proposed an improved method of detection which avoids this fallacy. Such a single-sideband system seems to be capable of the highest digit rates over good lines providing comprehensive line equalization is employed. Where the very highest speeds are less important than achieving low error rates under adverse conditions without undue complexity this method of modulation is inferior to phase and frequency modulation.

Those engineers acquainted with v.h.f. telephony techniques may be surprised that there should be any significant difference between p.m. and f.m. since they are usually treated as being essentially similar. The usual f.m. theory does in fact show that, with suitable modulation parameters, a carrier phase-modulated with a signal $f(t)$ is identical with one frequency-modulated with the differentiated signal $f'(t)$. When, as in telephony, the modulating signal is essentially sinusoidal its differential is also sinusoidal and the two signals have similar spectra. It is also usually possible at v.h.f. to make approximations in analysis by using the fact that the frequency of the carrier is much higher than that of the modulation and, in effect, this means that the detailed structure of the waveform can be neglected.

In data transmission the modulating waveform, at its simplest, is a square wave and, ideally at least, this involves an instantaneous jump of frequency or phase. A carrier modulated in frequency by a square wave is thus equivalent to a carrier appropriately modulated in phase by a triangular waveform. It is fairly obvious that the spectrum of this waveform will be considerably different from that of a carrier modulated in phase by a square wave. (The actual spectra are compared in Section 2.3.) Further, because we are dealing with carrier and modulating frequencies which are comparable, many of the mathematical simplifications of v.h.f. are not permissible and the detailed structure of the waveform is important.

The general treatment of frequency modulation is thus difficult in the data transmission context. It may be shown, however, that certain types of f.m. receivers have optimum performance when the difference between the two signal frequencies is equal to the digit rate (see Section 2.2) and all practical f.m. systems known to the authors use frequency separations fairly close to this value. If the frequency separation is in fact equal to the digit rate one digit contains exactly one more cycle than the other and it is possible to assume that the two signals are derived from locked oscillators. (A typical f.m. signal having this property, in which the digits consist of either $\frac{3}{4}$ or $1\frac{3}{4}$ cycles, is shown in Fig. 2(i).) The locked oscillator assumption makes relatively simple analysis possible and is therefore extremely convenient. It is usually apparent that the performance of f.m. systems which do not use locked oscillators approximates closely to the locked oscillator case though precise evaluation is not attempted. An advantage deriving from the use of locked oscillators in practical work is that stationary waveforms are obtainable.

To illustrate the points just discussed reference is made to the conditions which were regarded as typical in the choice of parameters for the practical work. These use two digit rates, 1500 and 750 bauds. For p.m. a carrier frequency of 1500 c/s is assumed at both digit rates while for f.m. the signal frequencies chosen are 1125 c/s and 2625 c/s at 1500 bauds, and 1125 c/s and 1875 c/s at 750 bauds.

2. Theoretical Comparison of P.M. and F.M.

When frequency modulation is used for data transmission a mark digit is normally represented by one frequency f_1 and a space digit by another frequency, f_2, where, as remarked, f_1-f_2 has invariably been chosen to be of the order of the digit rate. (We refer to the question of choice of frequencies again in Section 2.2.) If the bandwidth available does not extend above 3000 c/s then at speeds from about 1200 bauds upwards the digit represented by the lower frequency will contain less than one whole cycle.

When phase modulation is used a difficulty arises in defining the received phase. This could be achieved by establishing an absolute phase reference by special coding of the message at the receiver but should noise or other disturbance cause the reference to jump by half a cycle all subsequent digits would be in error. For this reason it is usual to avoid the problem by transmitting a mark as a phase reversal (i.e. 180 deg phase change) and a space as no change of phase. When differential working, as this method is called, is employed the receiver compares each digit with its predecessor; if a difference of phase is detected the receiver output is a mark and vice versa. While the exact effect of this depends to some extent on the detection method used and the point in the receiver at which the phase comparison is made, the ideal action is as follows. With differential working a digit which is wrongly sensed will cause a receiver error both for itself and for the subsequent digit with which its phase is compared. For isolated errors the error rate is therefore doubled by differential working. However certain patterns of interference may actually give rise to fewer errors when differential working is employed. For example where a non-differential system produces a continuous burst of errors the differential system will give only two errors, no matter how long the burst. Thus in impulsive noise the effect is less serious; this usually renders unreliable a group of, say, n adjacent digits and differential working merely extends this unreliability to $(n+1)$ digits. It may in fact be shown that if noise is of such an impulsive nature that it makes identification of n adjacent digits arbitrary, then ideally the error rate due to differential working is $(n+1)/n$ times greater than the error rate due to non-differential working.

The practical effect of differential working will be even less when error detection is employed. Digits are here divided into blocks of say D digits and the whole block is rejected if an error is detected. Impulsive noise which causes an n-digit error burst (i.e. with n digits between first and last error) will have $(n-1/)D$ probability of affecting two blocks $(n < D)$. If the error bursts occur independently, the average number of blocks in error per burst will be $1+(n-1)/D=(D+n-1)D$. Since the number of digits affected in the differential case will be increased by one the total number of blocks in error per burst will be raised to $(D+n)/D$. The increase in block error rate due to differential working is thus

$$\frac{D+n}{D+n-1} = 1+\frac{1}{D+n-1}$$

Typically if $D = 100$ and $n = 20$ the block error rate is increased by 0·84%.

The need for differential working must however be regarded as a disadvantage of phase modulation which may at very worst double the error rate. This appears to be the only respect in which f.m. commands superiority and it will emerge that in other respects p.m. has properties which substantially offset this disadvantage.

2.1. *Comparison of Ideal Systems*

It is useful to consider the performance of the two modulation methods under ideal conditions, that is the signal channel is assumed to be distortionless and the receiver is therefore fed with a signal which is perfect except for additive noise. In an ideal binary system the transmitter sends one of two possible signals. An ideal receiver has full knowledge of these two "expected" signals and their times of reception and calculates the probabilities that each received signal was due to the two transmitted signals, deciding in favour of the signal having the higher probability. Such a receiver will therefore have the lowest error rate that can be achieved with the signals chosen.

The theory of ideal receivers has been expounded by a number of authors. It will serve the present purpose to quote briefly results from a report by Becker and Lawton.[6]

Assume that both signals are equally probable and have equal energies per digit. If, in the presence of white noise, a signal $y(t)$ is received when either $x_0(t)$ or $x_1(t)$ was transmitted, then it may be shown that the probabilities that x_0 and x_1 were actually transmitted are

$$P_y(x_0) = \int_0^T y(t)x_0(t)\,dt \qquad \ldots(1)$$

and

$$P_y(x_1) = \int_0^T y(t)x_1(t)\,dt \qquad \ldots(2)$$

respectively, where the digit duration is from $t = 0$

to $t = T$. An ideal receiver computes these probabilities and selects the larger. This process is called correlation detection.

It may further be shown that the error probability, P_e, of such a signal is given by:

$$P_e = \frac{1}{2}\left[1-\mathrm{erf}\left\{\frac{E(1-\rho)}{2N_0}\right\}^{\frac{1}{2}}\right] \qquad \ldots(3)$$

where E = average signal energy per received digit,

N_0 = white noise power per unit bandwidth

and $$\rho = \frac{1}{E}\int_0^T x_0(t)x_1(t)\,\mathrm{d}t \qquad \ldots(4)$$

It will be noted that eqns. (1), (2) and (4) are relations involving the cross-correlations of y, x_0 and x_1.

In eqn. (3) the only parameter depending on the choice of modulation system is ρ. Equation (4) shows that this may vary between $+1$ (when the signals are identical and therefore indistinguishable) and -1 (when the two signals are as different as possible).

The case when $x_0(t) = -x_1(t)$ is an optimum and phase modulation with π-phase switching is in fact an application of this case.

If, in an f.m. system, the difference between the signalling frequencies is chosen to be equal to the digit rate, then $\rho = 0$. This value also applies when the frequency separation is high compared with the digit rate and each digit contains a large number of cycles; it is also a good approximation in all other cases of practical interest.

Pushman[10] has shown that where there is no phase lock in the transmitter and the signal therefore consists of one of two sine waves starting in the same unspecified phase then the cross-correlation coefficient cannot have an average value more negative than about $-0{\cdot}2$.

While in principle one can postulate a correlation detector which will take advantage of this negative value of ρ its design would be complex in the extreme; this will be seen by reference to Section 2.2 where the difficulties of implementing an f.m. correlation detector with locked signals are discussed. The more conventional f.m. receivers do not appear to be capable of using to advantage the negative value of ρ. It will thus be seen that $\rho = 0$ is a fair value to choose for f.m.

If the values -1 and 0 are successively substituted for ρ in eqn. (3) it is clear that to obtain the same probability of error E/N_0 must have twice the value for f.m. that it has for p.m. This is usually expressed by saying that phase modulation has a 3 dB advantage in signal to noise ratio over frequency modulation. The curves for eqn. (3) applied to p.m. and f.m. are shown in Fig. 1.

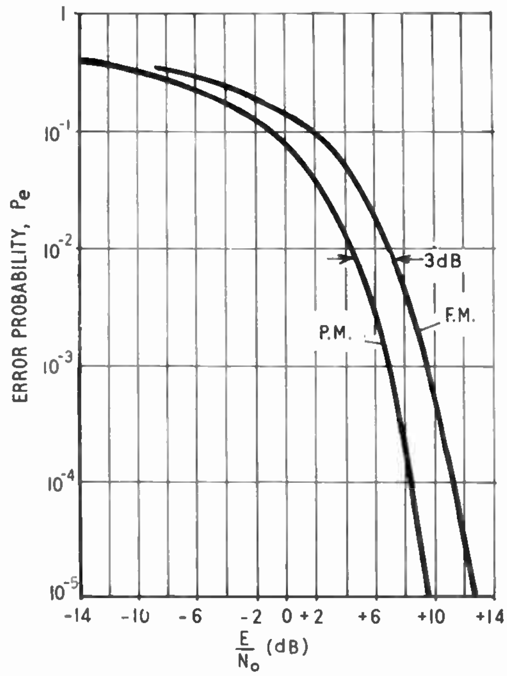


Fig. 1. Theoretical performance of ideal systems in white noise.

It must be stressed, however, that the mathematical computation neglects the effect of differential working and applies only to white noise. It also assumes that the bandwidths of signal and noise are not limited. Under such ideal conditions input filters will cause the correlation detector performance to deteriorate. In practice, of course, the ideal conditions are not achievable and it is probable that no serious error is involved in practical measurement if the signal spectrum shows that negligible signal power exists outside the noise bandwidth used. (An attempt to "normalize" signal/noise ratio from one bandwidth to another has been encountered which is based on the mistaken assumption that contributions from any part of the white noise spectrum affect the receiver performance equally.)

Clearly the cross-correlation coefficient (ρ) values show that there is more difference between the two p.m. signals than the two f.m. signals and in general therefore any arbitrary interference will probably cause more errors on f.m. than on p.m. Nevertheless it is possible to specify certain interfering signals for which the reverse is true.

An illustration of the fact that the 3 dB difference is due to the choice of signal and not the properties of white noise is obtained by considering any f.m. signal in which the signal frequencies and digit rate are locked. Assume for example a 1500 baud signal using frequencies of 1125 and 2625 c/s. A typical message waveform is shown in Fig. 2(i). A high-frequency digit may be regarded as a part of the continuous sine wave $\sin 2\pi\, 2625t$ and, similarly, the low-frequency digit is part of $\sin 2\pi\, 1125t$. It may be shown (see Section 2.3) that for this case the power density distribution of any binary symmetric message, including a random message, contains two line spectra at 1125 c/s and 2625 c/s which absorb half the signal power. These lines may therefore be represented by $\frac{1}{2}(\sin 2\pi\, 2625t + \sin 2\pi\, 1125t)$. Let us now subtract the line spectra from each digit in turn. The high-frequency digit now becomes $\frac{1}{2}(\sin 2\pi\, 2625t - \sin 2\pi\, 1125t)$ while the lower frequency digit becomes $-\frac{1}{2}(\sin 2\pi\, 2625t - \sin 2\pi\, 1125t)$. This process is illustrated graphically in Fig. 2(ii) and (iii), where it must be borne in mind that each digit follows a regular cycle and may occur in four possible phases. It will be noted that digits 2 and 6 are identical and digits 5, 7 and 8 are high-frequency digits in phases appropriate for comparison as binary alternatives with low-frequency digits 1, 3 and 4 respectively.

The significant feature is that after subtraction of the line components we are left with phase modulation using a signal given by $\frac{1}{2}(\sin 2\pi\, 2625t - \sin 2\pi\, 1125t)$. This may, for example, be seen in Fig. 2(iii) by comparing digits 1 and 5, 3 and 7, or 4 and 8, which are of the same waveform but in opposite phase and digits 2 and 6 which are identical. This phase-modulated signal has only half the original signal power and is therefore capable of the same white noise performance as the original signal. The inferior performance of f.m. may thus be attributed to wasteful components in the signal.

Pushman[10] gives an analysis on these lines for an f.m. signal in which the digit rate and signal frequencies are not locked. This analysis yields terms representing wasted power of a similar order of magnitude to the line spectra above.

There is thus a clear indication that an f.m. signal is composed of a p.m. signal plus components

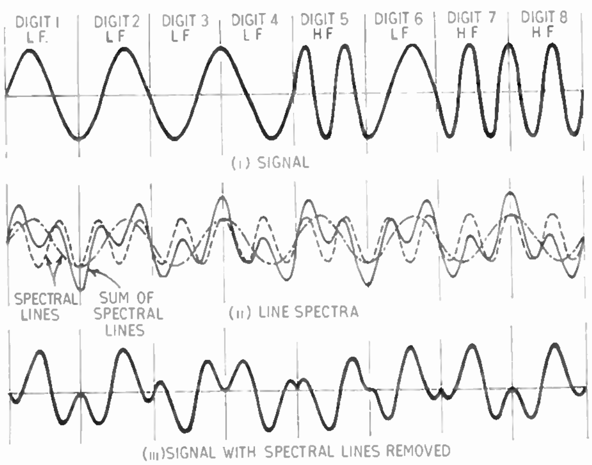


Fig. 2. Analysis of f.m. waveform.

containing a roughly equal amount of power which, because they are common to both f.m. digits, are valueless. The "hidden" p.m. signal uses a waveform dependent on the two f.m. frequencies. In the particular cases where interference can be shown to cause a lower error rate on f.m. than on p.m. the hidden p.m. signal also has this property. The reason for the relative performance thus lies in the choice of signal for the phase modulation process in the presence of the particular interference under consideration rather than in the modulation method itself. Normally a single frequency is chosen to form the basic p.m. signal because this simplifies the receiver but use of the compound waveform would have advantages if certain types of noise were prevalent.

The ideal receiver, which computes $P_y(x_0)$ and $P_y(x_1)$ according to eqns. (1) and (2), gives optimum performance only on white noise. According to Turin,[7] quoting an unpublished paper by R. M. Fano, a signal contaminated by non-white noise must first be passed through a "noise-whitening" filter which has a frequency response inverse to the spectrum of the noise. This signal is passed to the ideal receiver which evaluates eqns. (1) and (2). As the signal will have been distorted by the noise-whitening filter, this distorted signal must be used as x_0 and x_1 in these equations.

2.2. *Practical Application of Ideal Receiver Theory*

It is desirable at this stage to consider the extent to which it is possible to implement the principles of the ideal receiver in practice.

It is difficult to apply the noise-whitening filter theory to impulsive noise since reliable spectrum analyses of line noise are not available and, even if they were, the filter might well prove to be unrealizable. Since the background noise is widely distributed throughout the spectrum and the impulses occur only occasionally on a good line the spectrum is probably fairly uniform over the line bandwidth and, in addition, contains components outside that bandwidth due to pick-up close to the receiver. In the practical case the signal is band-limited and there is clearly no point in admitting noise components where no signal components exist. The best compromise at the receiver is to use a receiver input band-pass filter of minimal delay distortion which will reject components outside the expected signal bandwidth, followed by a correlation detector. The problem therefore of implementing the ideal receiver theory is reduced to that of achieving correlation detection with p.m. and f.m. signals.

A great deal of theory on optimum receivers uses the concept of the matched filter.[7] A filter is said to be matched to a signal $x_0(t)$ if its output on being fed with $y(t)$ is given by eqn. (1). There appears

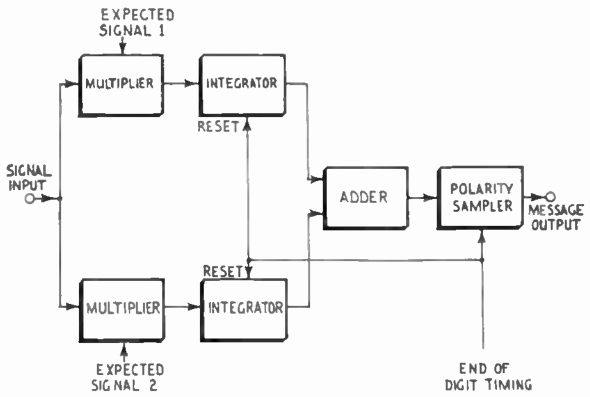


Fig. 3. The correlation detector.

however to be no passive circuit which is capable of performing this function for the signals under consideration. Even if a filter could be designed which was matched to a signal consisting of about one cycle of sinewave it is doubtful if it would still give a meaningful output when the signal was subjected to frequency translation which in effect varies arbitrarily the starting phase of the signal. Matching to a waveform envelope is an inferior alternative since it neglects information in the signal phase; it could therefore only be used on f.m. Envelope matching is, however, not feasible unless there are several cycles per digit and frequency changing would be necessary to make this possible (see later in this section).

True correlation detection for the data transmission case can therefore be achieved, if at all, only by use of multiplying and integrating circuits to perform the calculation of eqn. (1). A typical detector would therefore have the form of Fig. 3. It will be noticed that in addition to the input signal this kind of correlation detector requires inputs in the form of expected signals and these must be generated in some way in the receiver from the incoming signal. (We shall defer consideration of digit timing regeneration, which is common to all systems, to Section 2.8.)

When a phase-modulated signal is being detected the expected signals may easily be regenerated since the two signals are identical except for sign and one or other is always present. A simple form of regenerator full-wave rectifies the incoming signal to obtain a waveform which is the same for all messages, having its fundamental at twice the signal frequency. From this the expected signals, one the negative of the other, may easily be obtained. In practice with p.m. it is only necessary to regenerate one of the expected signals and to use one multiplier and one integrator. If a digit is received in one phase the output of the integrator will be positive and, if in the other phase, negative. The decision as to which phase is received is then made by sampling the polarity of the integrator output at the end of the digit period. A block diagram of such a receiver is shown in Fig. 4 with waveforms in Fig. 5. With differential working it is immaterial which phase of signal is actually regenerated since, after detection, the signals are examined only to see whether or not a change in phase has occurred compared with the previous digit. Furthermore, frequency translation in the channel is readily followed by the regenerator and correlation detection of the translated signal is achieved.

Because the carrier regenerator uses a filter its output, the expected signal, is substantially free from noise. The presence of this filter implies that the carrier frequency itself must, however, be reasonably constant at the transmitter and this filter also limits tolerance to rapid corrections of frequency translation drifts in the channel.

Digit timing information can conveniently be extracted from the waveform at the multiplier output. Although for clarity in Fig. 5 the waveforms are

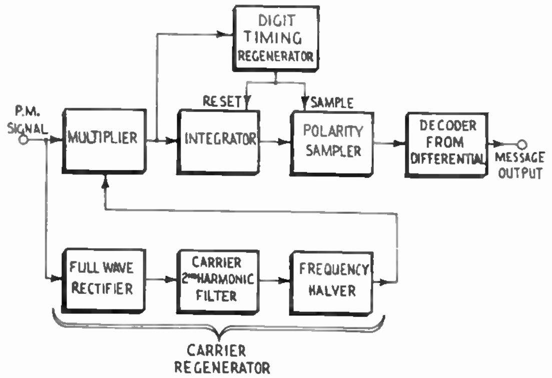


Fig. 4. Phase modulation receiver employing carrier regeneration.

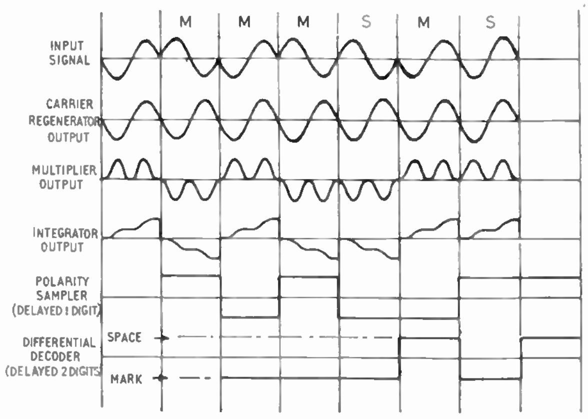


Fig. 5. Waveforms of Fig. 4 when digit rate and carrier frequency are equal.

shown with digit timing and carrier frequency locked, the carrier regeneration and digit timing circuits are quite independent and synchronism is not at all necessary. Thus, contrary to popular opinion, it is possible to use p.m. with a varying digit rate. (Tolerance to digit rate variations must ultimately depend on the flexibility of the digit timing circuits but this applies to all modulation systems.)

It is sometimes objected that this method of detection involves the recovery of digit timing in the receiver whereas other methods allow this to be done in the equipment supplied from the receiver. This is hardly a significant point since digit timing must be recovered eventually and if optimum performance is the aim the logical place for this to occur is in the receiver itself. However, if required, the integrator may be replaced by a low-pass filter and squarer and, at some loss of performance, a receiver output can be obtained without digit timing recovery.

It is also possible to detect phase modulation by comparing the phase of a digit as received with the phase of its predecessor providing that each digit consists approximately of a whole number of half-cycles. This is usually accomplished by using a line having 1 digit delay in the arrangement shown in Fig. 6. Waveforms for this circuit are shown in Fig. 7. Here both inputs to the multiplier are signals which may be noisy and this means some loss of performance. Apart from the noise on the reference this signal is virtually a correlation detector. However, large variations of digit rate and carrier frequency cannot be tolerated since the inputs to the multiplier lose the necessary relationship if this occurs. The circuit is, however, able to cope with very rapid corrections of frequency translation drifts in the channel. A further advantage is that decoding from the differential signal is implicit in the detection method and no special provision is needed for this. This can be seen by comparison of Figs. 5 and 7.

It will thus be seen that phase modulation lends itself readily to at least two practical arrangements which are inherently of the correlation detector type.

The application of correlation detection to frequency modulation is less straightforward and, in the view of the authors, is not a practical proposition.

Problems arise in the arrangement of Fig. 3 in regenerating the expected signals. The difficulty is eased somewhat if we assume that both expected signals can be represented by continuous sine waves. This implies that the transmitter signal frequencies are generated by the locked oscillator method since these frequencies must be separated by the digit rate if one signal is to contain exactly the one additional cycle per digit which is necessary to ensure smooth transitions from one digit waveform to the other.

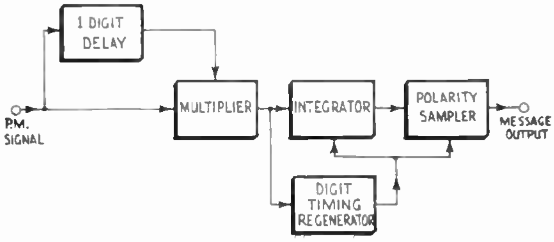


Fig. 6. P.m. delay line receiver.

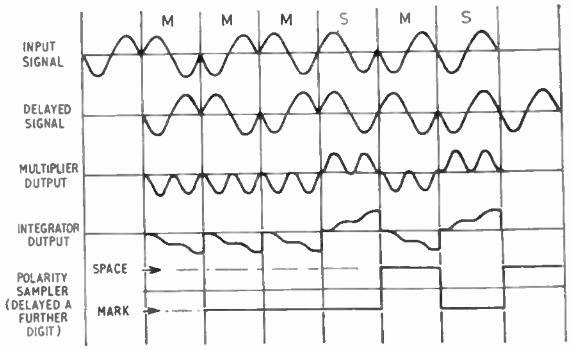


Fig. 7. Waveforms of p.m. delay line receiver.

The f.m. correlation detector must now regenerate from the incoming signals two sine waves which are controlled in frequency and phase by the corresponding digit signals. However, since one digit may be absent for an interval this is far from an easy problem to solve. The solution of relating harmonically the frequencies of the two digits so that each can be regenerated from the other as well as from itself cannot be applied because frequency translation in the channel destroys the harmonic relationship. It would, however, probably be feasible to beat the digit timing with the incoming signal to obtain the frequency of the digit not being received. Whilst apparently simple in principle such a scheme leads to considerable practical complexity. An f.m. correlation detector based on such an approach would be at least twice as complex as a p.m. correlation detector and its performance would inevitably fall further short of its ideal than the p.m. circuit, which, as we have already pointed out, has an intrinsic 3 dB advantage.

A form of correlation detector for f.m., based on a similar idea to the p.m. delay line receiver, is feasible but could only be used with certain signal frequencies. If both digits consist of a whole number of half-cycles and if the frequency difference is again equal to the digit rate, then the integrator in the arrangement of Fig. 6 will ideally give a zero output when a digit follows one at the other frequency. On the other hand, a non-zero output (positive if the digits consist of an even number of half-cycles, negative if odd) will

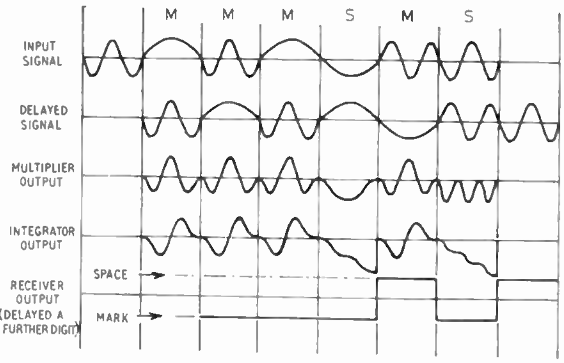


Fig. 8. Waveforms of f.m. delay line receiver.

occur when a digit follows one at the same frequency. The f.m. signal would therefore require to be coded differentially. Typical waveforms are shown in Fig. 8.

Another approach to the correlation detector using f.m. is to translate the incoming signal to an appreciably higher frequency so that there are several cycles per digit. It is still just as difficult to regenerate the actual expected signal but passive filters can now be used which will respond to the burst of several cycles to give an output whose envelope is virtually independent of the starting phase of the digit. Because such circuits ignore the phase content of the signal they do not make full use of the available signal information and appear to be at best envelope detectors. They do, however, conform to the normal concept of an f.m. receiver in that their output is derived essentially from information carried in the frequency only. The most common practice is to apply the translated signal to a discriminator which can, for example, consist of two circuits tuned approximately to the digit frequencies, the digit being identified from the circuit having the larger output. This arrangement is open to the objection that tuned circuits have storage properties and the output will to some extent be influenced by earlier digits in the message. Although, of course, circuits are usually designed to minimize such inter-digit interference, it is not in practice usually possible to eliminate it completely, nor is it likely that filters which give least inter-digit interference will also give best noise immunity.

Inter-digit interference can be better combated by use of the resonator principle. Here again it is desirable that the frequency separation should be equal to the digit rate. Two circuits of ideally infinite Q, tuned to either digit frequency are used, and these are fully quenched at the beginning of each digit. If it is fed with the frequency to which it is tuned the resonator output should have a linearly increasing envelope. When supplied with another frequency a beating effect takes place between the transient and steady state responses of the resonator and the resultant waveform has its first zero at a time equal to the reciprocal of the difference between the supplied frequency and the resonator natural frequency (hence the desirability of making the frequency separation equal to the digit rate).

It has been stated by Doelz and Heald[8] that the envelope sample obtained by the resonator method has amplitude given by

$$\{[\int_0^T S(t)W_1(t)\,dt]^2+[\int_0^T S(t)W_2(t)\,dt]^2\}^{\frac{1}{2}} \quad ..(5)$$

where $S(t)$ is the received signal representing a digit lasting from $t = 0$ to $t = T$, $W_1(t)$ is a sine wave of the same frequency as the expected signal (but of unrelated phase) and $W_2(t)$ is obtained by shifting $W_1(t)$ by 90 deg. The resemblance of eqn. (5) to eqn. (1) is immediately apparent. If $W_1(t)$ were in correct phase then the first integral would correspond exactly to the correlation detector equation and in fact on a noise-free signal the resonator and correlation detector outputs correspond for any phase of $W_1(t)$. Note, however, that the presence of uncorrelated noise is likely to produce as much output from each integration as occurs in the correlation detector. Whilst probably the best f.m. detector of a practical kind the resonator is thus significantly inferior to the f.m. correlation detector and an analysis for ideal performance in white noise shows that it is some 4 to 6 dB worse than the p.m. correlation detector[6] for signal/noise ratios of practical interest.

In order that the resonator may be quenched at the end of each digit period, digit timing must be available in the receiver. Unfortunately, there appears to be no obvious way of deriving this except by including some form of discriminator circuit and, for this reason, the resonator uses rather a large number of components.

Another f.m. detector which has found considerable practical application is the zero-crossing detector. It is difficult to relate this to the correlation detector, but as the only other circuit in common use it is appropriate to make a few comments at this point. The zero-crossing detector generates a pulse whenever the incoming signal passes through zero potential. A pulse frequency detector is then employed to count the number of cross-overs in each digit. In one simple form[9] this is an analogue counter consisting of an integrating capacitor which is discharged by each pulse but is charged linearly between pulses. The capacitor will thus charge to higher voltages on low-frequency digits than on high-frequency digits. If the capacitor voltage exceeds a chosen value during the digit period the digit is regarded as being of low frequency. We have in fact used in our tests an integrating arrangement which averages the frequency information over the digit period and is therefore slightly more efficient (see Section 3) but the principle is not changed.

The zero-crossing detector is difficult to analyse but approximate theory suggests that its performance is similar to that of the envelope detector. An interesting point in connection with the zero-crossing detector is that to provide one more zero crossing in the high-frequency digit than in the low-frequency digit the frequency separation must be at least half the digit rate. This would, however, leave little margin for error since, for example, one sine wave may contain, in the limit, three cross-overs whereas one and a half sine waves at the opposite limit also contain three cross-overs. The difference in the number of cross-overs is only certainly increased to two by increasing the frequency separation to a whole cycle per digit and this is therefore a reasonable choice.

It is significant to note therefore that the correlation, resonator and zero-crossing detectors all require a frequency separation equal to the digit rate for efficient operation. Where channel bandwidth is restricted it may be that the reduction in signal bandwidth obtainable from use of a smaller separation may more than offset the loss in detector efficiency but this can only be at a loss of performance and it is doubtful whether it is wise to use frequency modulation under such conditions.

The survey of practical receivers which has been presented indicates clearly that while a practical p.m. detector may be expected to approach the ideal p.m. performance closely, an f.m. detector is likely to fall appreciably below the ideal f.m. performance which is in any case intrinsically 3 dB inferior to its p.m. counterpart. These conclusions apply essentially to undistorted signals and the effect of channel impairments will now be considered.

2.3. *Bandwidth Requirements of P.M. and F.M. Systems*

It is fairly easy to show that a phase-modulated signal requires the same bandwidth as the corresponding amplitude modulated signal for, if there is added to any p.m. signal a continuous sinewave which is of carrier frequency and signal amplitude and in phase with one of the digit signals, it is found that the digits in this phase are doubled in amplitude whilst the digits in the opposite phase cancel and an a.m. signal for the same message results. These two signals differ only by the carrier frequency component and their spectra are otherwise identical.

Since less bandwidth is required for a.m. than for f.m. it will be clear that p.m. also requires less bandwidth than f.m.

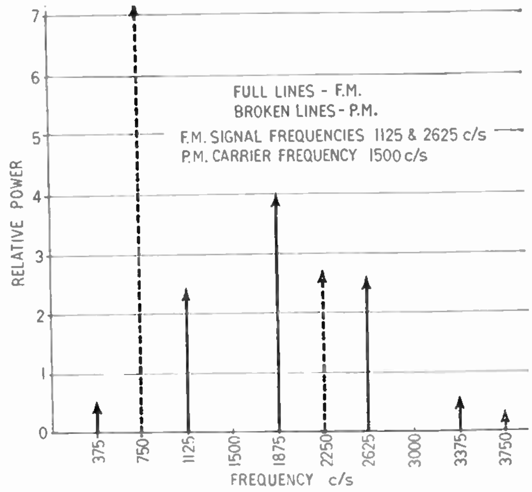


Fig. 9. Power spectrum of all-reversals message at 1500 bauds.

The relative bandwidth requirements can be analysed in more detail by considering the power spectra both for the worst case and for the average case. The spectral components are most widely distributed when there is a change of phase or frequency with each digit (the all-reversals message) and this message therefore constitutes a worst case. Since this is a repeated message, Fourier analysis can be applied and the spectrum consists of line components. The average case, on the other hand, requires an analysis of the random message. This subject has been investigated in some detail by Pushman[10] and some of his results are quoted here without proof.

Within the confines of this paper it is possible to discuss only a few typical cases. Figure 9 shows the all-reversals power spectra for a 1500 baud p.m. signal using a carrier frequency of 1500 c/s with switching at signal zeros (as in Fig. 7) and the corresponding spectra for a 1500 baud f.m. signal using signal frequencies of 1125 c/s and 2625 c/s locked to the digit rate and switching alternately at peaks and zeros (as in Fig. 2(i)). It will be noticed that the phase modulation signal has two strong components at 750 and 2250 c/s and a component at 3750 c/s which contains about 1·5% of the total power. Thus about 2% of the p.m. all-reversals signal power lies outside the line bandwidth. The f.m. signal has strong components at 1125 and 2625 c/s (the signal frequencies) containing half the signal power and a further strong component at 1875 c/s. Some 4½% of the total power is contained in each of two components at 375 and 3375 c/s close to or beyond the edge of the available bandwidth; other components are negligible. As has already been mentioned the components at the signal frequencies occur in all symmetric messages and

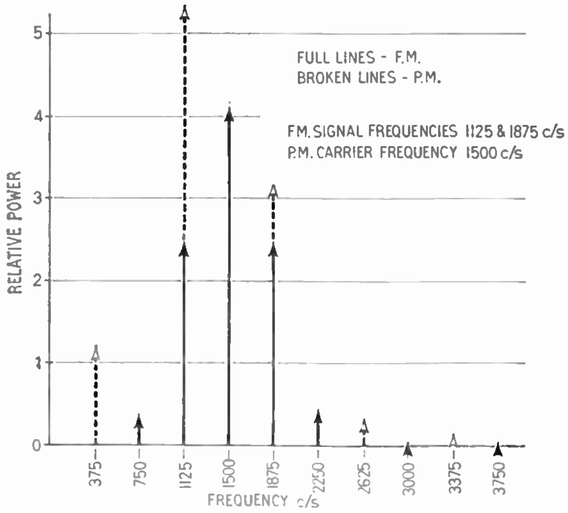


Fig. 10. Power spectrum of all-reversals message at 750 bauds.

therefore carry no information so it will be apparent that in this case some 18% of the useful f.m. signal power is close to or beyond the edge of the line bandwidth.

The all-reversals spectrum at 750 bauds using for p.m. a 1500 c/s carrier frequency switching at signal zeros and for f.m. signal frequencies of 1125 and 1875 c/s switching at the signal peaks is shown in Fig. 10. In the p.m. case the two largest components are at 1125 and 1875 c/s and these, with a further small component at 2625 c/s, are likely to be within the available bandwidth. Components near or beyond the band-edge at 375 c/s and 3375 c/s carry 11 and 0·6% of the signal power respectively. In the f.m. case the "wasted" components occur at 1125 and 1875 c/s and a strong useful component occurs at 1500 c/s with others at 750 and 2250 c/s each containing about 4% of the total power. In this case f.m. appears to have the advantage, due no doubt to the favourable choice of switching instant which keeps the waveform and its first differential coefficient continuous and each digit without a d.c. component (advantages which can be maintained only if locked oscillators are employed).

The random spectra for these cases are shown in Figs. 11 and 12. It will be noted that again the f.m. signals have line spectra using half the total power at the signal frequencies.

At both speeds p.m. has the narrower random spectrum. At 1500 bauds the bandwidth for p.m. and f.m. at half peak power are 1350 and 1850 c/s respectively, and the skirts of the f.m. spectrum are also less favourably shaped. The corresponding bandwidths at 750 bauds are 665 and 870 c/s.

At the lower speeds the p.m. advantage is not very marked. The smoother transitions available with f.m. largely offset the fact that two fairly widely separated frequencies have to be used. However, although both modulation methods require a bandwidth which increases linearly with speed the constant of proportionality is greater for f.m. than for p.m. and at the higher speeds p.m. requires a significantly narrower bandwidth than does f.m.

In practice a truly random message would not be transmitted since this could occasionally give rise to long periods without reversals and the receiver would be in danger of losing the digit timing. The simplest way to avoid this difficulty is to make every nth digit a reversal. Pushman[10] has shown how to derive the spectrum when, in an otherwise random message, every nth digit is a reversal. The spectrum so derived for $n = 10$ is not greatly dissimilar from the truly random case.

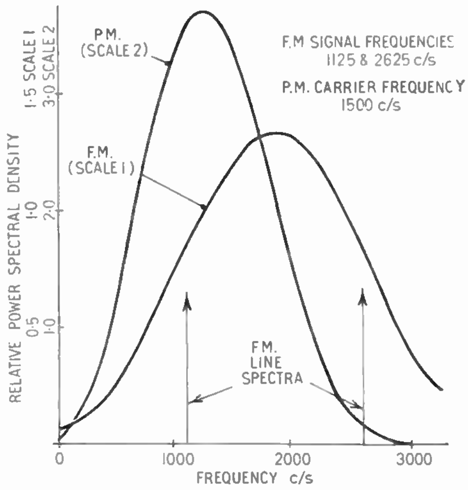


Fig. 11. Random message spectral density distribution, 1500 bauds.

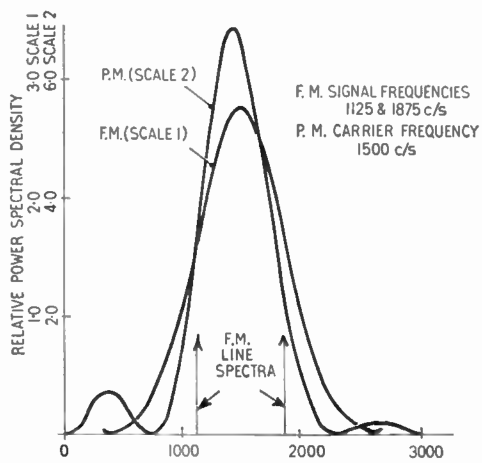


Fig. 12. Random message spectral density distribution, 750 bauds.

2.4. *Amplitude and Delay Distortion*

The discussion of the last Section means inevitably that, especially at the higher digit rates, f.m. will be more susceptible than p.m. to the effects of amplitude and delay distortion since these are always worst towards the band-edges.

Within limits it is of course possible to equalize for these distortions but invariably the wider-band signal will require a more complex network and more exacting component tolerances.

It has become common practice in f.m. receivers to include correction for bias distortion. This form of distortion arises when the channel treats the two kinds of digit differently and is frequently due to amplitude and delay distortion. Bias distortion correction which must be adjusted to suit the line in use, seeks to cancel one distortion with another and is, at best, a compromise. Because of the inherent symmetry of its signal, however, p.m. does not suffer from bias distortion.

In any optimum binary communication system both signals should suffer identical impairments and so be liable to the same error rates. The lack of bias distortion in p.m. is a particular example of this general property which is possessed only by the p.m. signal.

It gives an important practical advantage in avoiding the danger encountered in f.m. design that measures adopted in the receiver to improve response to one of the digits can affect the other adversely. Because of this it is necessary to carry out an elaborate series of measurements to be sure that an f.m. receiver is set up for optimum performance.

A technique known as "smearing" has been proposed[11] to combat impulsive noise. This involves delay distorting the signal before transmission and correcting for this distortion by equalization in the receiver. Phase modulation is more adaptable to this technique because the networks required have to be accurate over a smaller bandwidth and are therefore simpler to design and construct.

2.5. *Frequency Translation*

It has already been remarked (Section 2.2) that the presence of the frequency translation effect precludes the use of a comparatively simple f.m. correlation detector based on a simple harmonic relationship between the two signal frequencies.

Experiments have shown that a p.m. correlation detector using the carrier regeneration method gives a performance no worse than would be expected due solely to the bandwidth limitation introduced by the translation equipment. It may therefore be inferred that this effect has no serious consequences. The simple circuit tested was able to tolerate translations up to 18 c/s and there appears to be no reason why this range should not be increased if required.

However, in discussing practical receivers in Section 2.2 it was pointed out that rapid frequency corrections can adversely affect carrier regeneration receivers. Such an effect does at present occur on some G.P.O. lines but it is understood that it is intended to eliminate this. The delay line method of p.m. detection is not greatly susceptible to this effect and is also capable of near optimum performance but the greater flexibility of the carrier regeneration receiver is certainly worth retaining wherever possible.

2.6. *Effects of Line Noise*

This is not a subject which can be dealt with at length here and only the points which bear on the choice of modulation system are mentioned.

As already indicated in Section 2.1, the 3 dB advantage of p.m. in white noise is essentially due to a property of the signals and is likely to apply in other forms of interference also. Nevertheless it is possible to specify noise which will produce errors on p.m. but not on f.m.

Study of the kinds of interference typically encountered on telephone lines shows that some of it is of this character. It transpires, however, that where f.m. has an advantage of this kind it is always comparatively small and applies over only a rather limited range of noise frequencies; it would almost certainly be neutralized by the greater distortion suffered by the wider band f.m. signal. On the other hand, where p.m. has the advantage this is frequently marked and could affect error rates considerably.

In the practical tests described in Section 3 a recorded sample of line noise was used which was as favourable to f.m. as could be obtained. With idealized receivers and no bandwidth limitation the performance for f.m. was in fact found at high noise levels to be somewhat better than for p.m. but with practical signals this advantage disappeared and use of practical receivers further increased the inferiority of frequency modulation.

2.7. *Change of Digit Rate*

A small but worthwhile advantage in favour of p.m., if carrier regeneration is employed in the receiver, occurs where different digit rates are to be used. It is usually possible to retain the same carrier frequency for phase modulation at the various speeds without sacrificing seriously the advantages of reduced data rate. It is therefore possible to provide change of speed facilities with few additions to the circuit. On the other hand, if an f.m. receiver is to derive full advantage from reduced digit rates it is necessary to modify one or both signal frequencies accordingly since the frequency separation appropriate to a high digit rate would lead to an unduly large signal bandwidth if used at lower speeds. This involves greater circuit complexity in both transmitter and receiver.

2.8. *Digit Timing Extraction*

Whatever system of detection is used some form of digit timing must be extracted either in the receiver or in the equipment which uses the receiver output. Since having this information available usually makes it possible to improve performance in the receiver, it is logical to regard digit timing extraction as a receiver function.

In all methods of detection an advantage is gained if the detected signal is effectively summed over the digit period so that the whole of the available information is utilized, a process which requires digit timing. The integrator in the correlation detector performs this function and it is equally desirable that the same principle be applied to other receivers.

Digit timing information, which is of course contained only in message reversals, is usually best extracted from the signal which is being supplied to the integrator. (As mentioned in Section 2.2 the f.m. resonator detector requires a somewhat different arrangement.) The greater efficiency of the p.m. signal normally means that the useful signal to noise ratio at this point is greater for p.m. than for f.m. and to this extent better digit timing regeneration should be possible. Apart from this factor there seems to be little to choose between the various receivers except for the f.m. zero-crossing detector which is likely to have additional jitter due to the variation in zero-crossing times caused by signals which do not consist of a whole number of half-cycles and by frequency translation.[12]

Providing the transmitter digit rate is constant, sophisticated clock circuits employing quartz crystals can be used to average timing over a large number of reversals and so make these circuits virtually immune to noise and disconnections over periods of several seconds. At the other extreme it is possible to obtain passable performance from a relaxation oscillator which has a slightly longer period than the true digit time and which is restored to correct phase by a trigger pulse derived from each receiver reversal. This arrangement has no averaging property and it is the usual practice to provide this by means of a tuned circuit which can extrapolate between reversals and also limit the amount of phase correction due to individual reversals, thereby giving a measure of immunity to noise and jitter.

2.9. *Relative Equipment Complexities*

It is difficult to assess relative equipment requirements since some circuits are desirable rather than essential and could be omitted where economy is the over-riding consideration. There is also the point that the same function can be performed in different ways each having their own advantages.

To give a simple comparison an estimate has been made of the number of transistors which would be required to give a high performance circuit for use at 750 and 1500 bauds. Each transistor will on average be associated with five or six small components. Inductors have been listed separately since they are the bulkiest and most expensive items in general use. Where alternative circuits are possible the most economical arrangement has been chosen; this assumes that the cost of two transistors and associated circuits is approximately the same as that of one inductor.

2.9.1. Senders

Several types of f.m. sender are available. The correlation detector if attempted would require locked oscillators in the sender. The f.m. delay line detector requires reasonable synchronism and locked oscillators are probably desirable. With a whole number of half sine waves per digit, however, it is easier to generate locked square waves at the signal frequency and to obtain the signal from this by filtering.

An unlocked f.m. signal can also be produced by filtering a rectangular wave obtained from a free-running multivibrator whose base leak is taken to a potential which is controlled by the message. This is the most economical form of f.m. sender but the signal frequencies obtained are not precisely defined.

The frequency variation required can also be obtained by incorporating a variable reactance device in the tuned circuit of an L-C oscillator. If the oscillator operates at line frequencies a large reactance variation is needed and it is difficult both to control the frequency and to ensure that the two digit frequencies occur at the same amplitude. The amplitude control can be improved at the expense of the frequency accuracy by operating the variable reactance oscillator at a higher frequency and by changing down to signal frequency subsequently.

Whilst variants are possible on the p.m. sender also, the most satisfactory device, using a sine wave source which is switched in a diode ring modulator by a bistable which changes state at each message mark to produce differential coding of the signal, is quite economical and is the obvious choice.

The estimated complexities of these senders are given in Table 1.

2.9.2. Receivers

The general principles on which the receivers are based are discussed in Section 2.2. Some variations in detail are possible, however.

Table 1

Relative Sender Complexities

	Transistors	Inductors
1. P.M.	8	4
2. F.M., Locked Oscillators	17	3
3. F.M., Variable Reactance	6	2
4. F.M., Variable Reactance and Frequency Change	8	4
5. F.M., Multivibrator and Filter	6	2
6. F.M., Locked Square Wave and Filter	12	7

The multiplier of the correlation detectors may be a true multiplier or a switched multiplier (the latter gives an output equivalent to multiplying the incoming signal by a square wave at expected signal frequency).

Automatic gain control, although advantageous, is not strictly necessary in an f.m. zero-crossing detector, and p.m. receivers have also been devised in which the signal is passed through a limiting amplifier rather than an a.g.c. circuit.

Digit timing circuits may vary in complexity from about 5 transistors and 1 inductor to 80 transistors and 1 inductor with, in the latter case, a quartz crystal and small motor to give refined frequency control.

Approximate complexities (excluding digit timing circuits) are listed in Table 2. The p.m. circuits and f.m. delay line receiver are assumed to employ differential working on the line but to give a decoded output.

It will be seen that the f.m. zero-crossing detector, especially in its least elaborate form, requires rather less equipment than the other circuits but that, in terms of complexity, there is not a great deal to choose between the other circuits.

Table 2

Relative Receiver Complexities (excluding digit timing circuits)

	Transistors	Inductors
1. P.M. Correlation Detector, True Multiplier	44	7
2. P.M. Correlation Detector, Switched Multiplier	40	7
3. P.M. Delay Line Detector	20	23
4. F.M. Zero-Crossing Detector with A.G.C.	38	6
5. F.M. Zero-Crossing Detector, simplified and without A.G.C.	32	6
6. F.M. Frequency Translation Discriminator	32	10
7. F.M. Frequency Translation and Resonator	40	11
8. F.M. Delay Line	22	23

If performance and complexity are both taken into account p.m. has a clear advantage where high performance is required.

3. Measured Error Performances on Practical Receivers

In this Section a brief account is given of experiments which were carried out to determine the relative performances of the various receivers under practical conditions.

To accomplish this it is of first importance that the measurements be taken under truly comparable conditions. It is worth remarking here that the results reported by various workers in the literature are not suitable for comparison purposes since even when the line used is specified in detail the type of interference is never defined in such a way that comparisons can be made.

Results have also been published giving performances in white noise. In these cases it is usual to find that before being applied to the receiver the signal and noise have been passed through a band-pass filter, the signal/noise ratio being measured at the filter output. Such measurements are meaningless unless the filter response is provided. A practical filter having the necessary sharp cut-off will almost certainly have some ripple in its amplitude characteristics and measurements show that the exact location of the response peaks is often much more important than the actual filter bandwidth. This arises because error rate is so steep a function of white noise level (see Fig. 1) and because noise components close to the signal spectrum peak are much more likely to cause errors than are components in other parts of the spectrum.

In order to provide truly comparable conditions, therefore, we have used an artificial (and therefore noise-free) line and have added to the signal at its output the same recorded line noise for each test. (The noise having been recorded at a line output, it seemed undesirable to pass it through the artificial line since this might well have excluded typical components.)

Two different artificial lines were in fact used and their characteristics are given in Fig. 13. As the main interest was in performance under adverse conditions no equalization was employed.

The sample of noise used in the tests reported in this paper was recorded on a long-distance link between Christchurch and The Hague and was of selected noisy portions of a much longer recording. (Because the noise was selected in this way error rates measured are much higher than should occur in practice.) Whilst at the outset of the study no preference was felt for either modulation method,

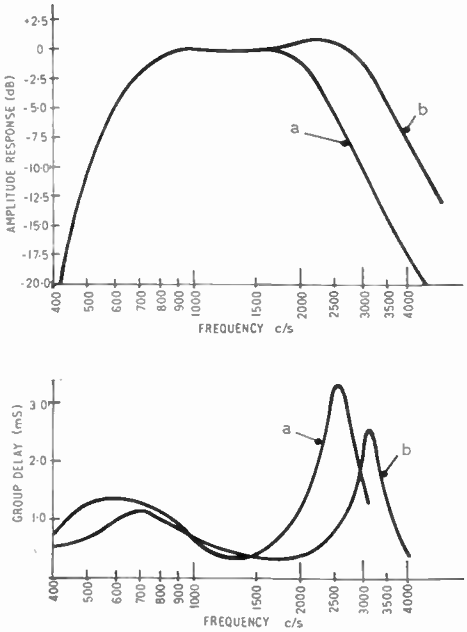


Fig. 13. Artificial line characteristics.

the theoretical work discussed in earlier sections had pointed to the superiority of p.m. The noise recording chosen was particularly suitable for establishing this point, since it provided many samples of large amplitude interference, which were of the kinds expected to favour f.m. (see Section 2.6). The results quoted are therefore as unfavourable to p.m. as are ever likely to be encountered in practice.

Before beginning the tests it was necessary to optimize the receivers. This was relatively easy for the p.m. carrier regeneration correlation detector since its performance in white noise with ideal signals should approach that of the ideal detector shown in Fig. 1. To provide a reference of the best that could be expected from f.m. a correlation detector was also used and again Fig. 1 provides a criterion of performance. In the f.m. receiver no attempt was made to regenerate the expected signals; these were supplied instead direct from the transmitter and were adjusted in phase to give best results.

A fairly sensitive check that a correlation detector is approaching ideal performance is obtained by ascertaining that with undistorted signals the error rates for various kinds of message are the same. The p.m. and f.m. correlation detectors were developed until this condition was achieved. The performances in white noise were then as shown in Fig. 14. It will be seen that when allowance is made for differential working these tally closely with the theoretical curves of Fig. 1 and in particular that they differ by approximately 3 dB in E/N_0. It must, however, be stated that a slight discrepancy may exist since the white noise source used was of finite bandwidth (about 35 kc/s) and the calculation of noise density was not free from error. However, the relative measurements were probably accurate to better than $\pm 0{\cdot}5$ dB and it is believed that the absolute accuracy was better than ± 1 dB.

Optimizing the other receivers was more difficult since no criterion of performance was available. However, in all cases where unknown variables were involved (as for example in deciding optimum Q of frequency translation f.m. discriminator circuits) tests were conducted to determine the best settings. The principle of integrating digit information over the digit period was adapted to all receivers and, for this reason, the results given will probably compare favourably with those obtainable from many commercial receivers of the various types. Delay line detectors were not tested as theory shows these to be not unduly inferior to their respective correlation detectors.

Tests on the impulsive noise recording with input filters showed that some improvement in performance could be obtained by subjecting the signal with noise to bandwidth limitation. Tests with amplitude limiters both before and after the filters usually produced a deterioration in performance, however. This was attributable to the fact that where a noise impulse exceeds the limiting level signal information is inevitably destroyed by the limiting action. (The desirability of using a linear detector is, in part, due to a similar effect.) The practical results quoted later in this Section were obtained with appropriate input filters in circuit.

For simplicity and reliability of testing, digit timing was not regenerated but was provided direct, in variable phase, from the transmitter. As explained in Section 2.8, this timing can be obtained virtually noise-free if desired and as this is much more a matter

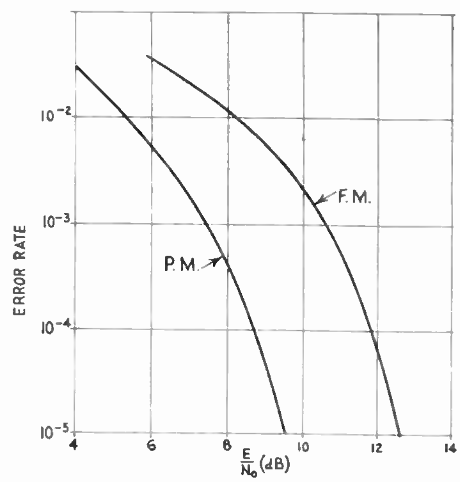


Fig. 14. Measured performance of correlation detectors with ideal signals in white noise, 1500 bauds.

of the complexity which can be tolerated than of the modulation method used it was deemed preferable to simulate perfect digit timing recovery.

There can be no doubt that tests reported by some other workers have been of only limited value because the messages transmitted were of undue simplicity. The message generator used in our tests was capable of producing an n-digit fixed message or a cycle of 2^n successive all different n-digit messages for all values of n from 3 to 20. The cycle was so constructed that if each of the $n2^n$ digits in a complete cycle was regarded in turn as beginning an n-digit message then each particular n-digit message occurred n times per cycle. Inevitably some sequences greater than n-digits in length were present also. The need for a sufficiently long and varied message is created in part by the presence of delay distortion, which introduces inter-digit interference and may make it impossible to receive without error certain combinations of digit even in the absence of noise. By suitably adjusting detecting circuits it was usually found to be possible to receive any fixed noise-free message without error. It was, however, not possible to tell in a fixed message whether one was countering the presence of delay distortion for the given message whilst making it worse for other messages. For all the practical tests the message generator was operated in the cycling mode with $n = 10$.

The set-up for the tests which simulated practical conditions is shown in Fig. 15. The tests were carried out at 750 and 1500 bauds. On p.m., differential working was employed and the carrier frequency was 1500 c/s whilst on f.m., which did not require differential working, signal frequencies of 1125 and 2625 c/s at the higher speed and 1125 and 1875 c/s at the lower speed were used. It will be noted that these f.m. frequencies are quite symmetrically disposed in the wide and narrow bandwidth lines respectively whilst the p.m. carrier frequency is quite centrally placed in the narrow band line response.

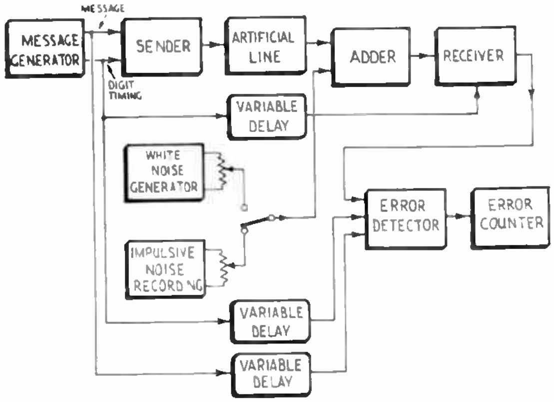


Fig. 15. Block diagram of arrangement of error performance tests.

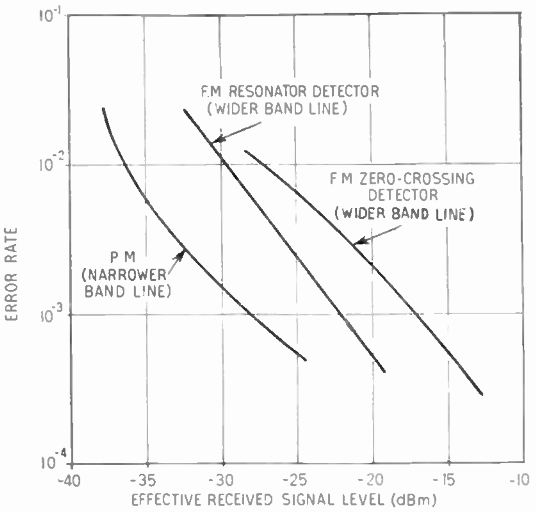


Fig. 16. Error rates at 1500 bauds.

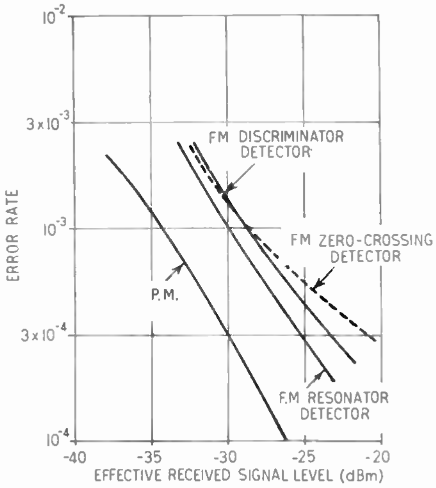


Fig. 17. Error rates at 750 bauds. (All tests use filters and narrower band line.)

The noise level was related to a reference tone on the noise recording which represented a received level of -38 dBm. It was convenient to vary the noise playback level relative to the signal and the effective received signal level was determined in relation to the reference tone.

The results obtained with the various receivers at 1500 bauds are shown in Fig. 16. It was not possible to receive f.m. without errors at 1500 bauds through the wider bandwidth line, even in the absence of noise, with the discriminator detector, though the resonator and zero-crossing detectors could be set up to operate in this way. It was, however, possible to receive 1500 baud p.m. signals with the narrower band line and it will be observed that the results obtained were some 7 dB better than with the resonator detector despite the fact that some 900 c/s less bandwidth was required. At this digit rate the zero-crossing detector was from $2\frac{1}{2}$ to 5 dB worse than the resonator detector.

At 750 bauds it was possible to use the narrower bandwidth line for both systems and, as would be expected, the comparison, though still unfavourable to f.m., is less marked. The results are shown in Fig. 17 from which it can be seen that at this speed p.m. is of the order of 5 dB better in performance than the resonator detector and a further 1 dB or more better than the more usual type of f.m. detectors.

4. Conclusions

It has been shown that phase modulation is considerably superior to frequency modulation in the data transmission application. The factors contributing to the p.m. advantage are:

(1) The f.m. signal has line spectra at the signalling frequencies representing about half the signal power. These components do not help to distinguish between messages and represent wasted power. The p.m. signal contains no component of this kind. Phase modulation therefore has an inherent 3 dB advantage in signal to noise ratio.

(2) The practical p.m. receiver works on correlation detector principles and therefore has performance close to the p.m. optimum. Practical f.m. receivers are at best envelope detectors and even with ideal signals fall 4 to 6 dB below the p.m. optimum.

(3) The f.m. signal requires a wider bandwidth than the p.m. signal, the difference becoming more pronounced as the digit rate is increased. Frequency modulation is therefore more susceptible to line distortion than p.m. and under practical conditions f.m. compares even less favourably than under ideal conditions.

(4) Change of signalling rate to reduce signal bandwidth may be more easily accomplished on p.m. than on f.m.

On the other hand, p.m. has the disadvantage that differential working is required. Compared with the ideal, this can lead to a doubling of the error rate in white noise, though in impulsive noise this margin is reduced and in systems having error detection it is of little importance. Synchronous operation (i.e. locking of signal frequencies and digit rates) is not essential on either modulation method.

The equipment complexity of good quality receivers is much the same for either modulation method. Where economy is of compelling importance an f.m. receiver may, at the price of a further deterioration in performance, be made more cheaply.

The resonator detector gives the best performance of the f.m. receivers. Using this as the f.m. detector, performance tests under practical conditions show that at 1500 bauds p.m. has an advantage over f.m. of about 7 dB in signal/noise ratio when passed through a channel some 900 c/s narrower in bandwidth, while at 750 bauds, using the same channels, the p.m. advantage is about 5 dB.

5. Acknowledgments

The authors would like to acknowledge the help given in valuable discussions by Mr. H. Gates of S.R.D.E. and Messrs. W. Renwick, H. J. Pushman and H. J. Faulkner of The Plessey Company Limited.

Thanks are also due to The Plessey Company Limited for permission to publish this paper.

6. References

1. A. A. Alexander, R. M. Gryb and D. W. Nast, "Capabilities of the telephone network for data transmission", *Bell Syst. Tech. J.*, **39**, No. 3, pp. 431–76, May 1960.
2. R. G. Enticknap, "Errors in data transmission systems", *Trans. Inst. Radio Engrs (Communication Systems)*, **CS-9**, No. 1, pp. 15–20, March 1961.
3. A. W. Horton, Jr. and H. E. Vaughan, "Transmission of digital information over telephone circuits", *Bell Syst. Tech. J.*, **34**, No. 3, pp. 511–28, May 1955.
4. J. L. Hollis, "Digital data fundamentals and the two levels vestigial-sideband system for voice bandwidth circuits", *I.R.E. Wescon Convention Record*, **4**, Part 5, pp. 132–45, 1960.
5. A. J. H. Oxford, British Provisional Patent Application No. 43886/61, 7th December 1961.
6. H. D. Becker and J. G. Lawton, "Theoretical Comparison of Binary Data Transmission Systems", Cornell Aeronautical Laboratory Inc., Report No. CA-1172-S-1, May 1958.
7. G. L. Turin, "An introduction to matched filters", *Trans. Inst. Radio Engrs (Information Theory)*, **IT-6**, No. 3, pp. 311–29, June 1960.
8. M. L. Doelz and E. T. Heald, "A predicted wave radio teletype system", *I.R.E. Convention Record*, **3**, Pt. 8, pp. 63–9, 1954.
9. L. A. Weber, "A frequency modulation digital subset for data transmission over telephone lines", *Communications and Electronics*, No. 40, pp. 867–72, January 1959.
10. H. J. Pushman, "Spectral density distributions of signals for binary data transmission", *J. Brit.I.R.E.* (To be published.)
11. R. Wainwright, "Overcoming Impulse Noise Interference in Narrow Band Data Communication Systems by a Sophisticated Filter Technique", Rixon Engineering Bulletin No. 70, July 1960. (Presented at the Rome Utica I.R.E. Conference in October 1960.)
12. I. C. Hinckfuss, "A Digital Data Transmission System, Part 1", Report EID6, Weapons Research Establishment, Australian Defence Scientific Service, Department of Supply, January 1960.

Manuscript first received by the Institution on 6th December 1961 *and in final form on 24th April* 1962. (*Paper No.* 737/C39.)

U.D.C. 62.007

Human Engineering

By

S. G. RAMSAY, B.Sc.†

Presented at a meeting of the North Eastern Section in Newcastle-upon-Tyne on 14th December 1960.

Summary: After defining the nature and scope of the field and indicating the roles of those specialists who contribute towards it, a short historical background is described. Some of the special difficulties encountered in experimentation with human beings are briefly covered, followed by a description of some typical problem areas such as design and layout of displays, design of equipment and controls, illumination, noise, etc. Finally, consideration is given to the practical and theoretical implications of findings in this field for both human behaviour in general and for industrial and business situations in particular.

1. Introduction

Although the field of inquiry which is to be reviewed in this paper is adequately distinguishable by using the term " Human Engineering ", it is important to realize that other terms are also currently in use, e.g. "Ergonomics", "Engineering Psychology", "Bio-mechanics".

All these terms refer to the area of investigation popularly known as "fitting the job to the man" but which has been more accurately described as the adaptation of work equipment, work space and work environment for optimal human use. To be more specific, the problem is to match human behaviour and the work environment, giving concentrated attention to the physical, physiological and psychological characteristics of that environment, with the aims not only of increasing output but of reducing stress, discomfort, fatigue, etc., and promoting industrial health and welfare.

Now, clearly, there is nothing very novel about such aspirations for even primitive man adapted his coracle so that he could sit more comfortably in it, and the development of medieval weapons shows that attention was paid to their design in relation to comfort and use. In the second half of the fourteenth century, for example, the swords and daggers of knights were frequently fastened not only at the waist but also by means of a special chain hanging from the chest. This was intended to prevent the weapon from getting lost if it was knocked out of the hand during battle. A study of the development of other weapons and parts of armour shows that they were fashioned not only to suit changing battle conditions and methods of fighting, but also, sometimes, to provide the maximal comfort for user or wearer. Throughout man's cultural and technological development, attention has been paid to the kinds of problems and their solutions with which we are concerned.

† Department of Psychology, University of Durham.

Where the difference lies, of course, between historically recent and historically distant approaches is in the former's emphasis upon a systematic, scientific approach. Human engineering is concerned with the relation between man and tool, man and machine, or man and environment.

Since man himself is a central figure in this kind of research and its applications, it becomes immediately obvious that, despite its title, human engineering is far from being the exclusive concern of the engineer alone. Problems in human engineering are liable to require information about body structure and function, levels of human energy, rates of metabolism and other physiological processes—as well as involving data regarding the psychological characteristics of the worker.

Properly, the main object of study in human engineering is neither man nor machine taken individually and separately, but the total man-machine system and its relation to efficiency. Just as a badly-designed machine is liable to turn out defective products, so is the bad operator or the uncomfortable and fatigued worker liable to show inefficiency as well as dissatisfaction in his work with machine, instrument or tool. Of course one must not regard human engineering as mutually incompatible with systems of guidance, selection, placement and training. Efficient procedures of these kinds also aim to reduce the number of inefficient and dissatisfied workers. Very often, however, situations arise in which these methods can do no more, e.g. when the labour demand is small, or when personnel cannot be further "adapted" or made more flexible. In such situations the approach via human engineering can be especially valuable.

Another consequence of the human organism being a central focus in human engineering is that we find all sorts of scientists and practitioners involved in such research and application. We find chemists, physicists, anthropologists, anatomists, physiologists, biochemists, statisticians, psychologists, etc., as well as many types of engineers, managers and industrial administrators. The latter are ultimately called upon to make the final decisions regarding the design and setting up of the work equipment and environment, and their discharging of responsibilities has a direct bearing on work performance and on human welfare in general. Errors can be costly in both human and economic terms but not all of these errors are necessary, let alone inevitable. Research has built up a promising accumulation of information; but in many situations the engineer or designer may still have to fall back upon his best judgment. The aim of human engineering is to reduce the number of situations in which the costly "rule of thumb" or guesswork are employed as the prime principles.

Of course it is not essential that a person who might call himself a student of human engineering be on the spot to advise. It is perfectly reasonable to use the findings of relevant, scientific research provided the engineer, manager or designer is able correctly to interpret these findings. In addition, there is no reason why preliminary trials of the performance of human subjects on a new task, or on a familiar task with new equipment, cannot be carried out. (It is perhaps worth pointing out that at the present time there is no separate professional practitioner known as a human engineer or ergonomist, who provides expert advice by consultation—not, anyway, in the same sense as does, say, the medical specialist. Currently, however, attempts are being made to look into the organization, training and status of personnel doing research in human engineering.)

In summary, then, the goals of human engineering are those of increasing man's efficiency and economy at work; an attempt to attain an improvement in performance at work coupled with an improvement in human welfare in general. In many cases the problem is one of altering the environment not to improve the performance of the total system (man plus machine), but to improve the comfort, safety, efficiency etc., of the man who is only part of that system.

2. Historical Development

Although the pre-scientific origins of human engineering can be found deeply embedded in man's history, as a more or less separate branch of occupational psychology and of engineering, and as a systematic, scientific approach to industrial problems, its beginnings are relatively recent. Human engineering really separated itself from the rest of industrial psychology only during and since the Second World War and as yet it is difficult to bring its boundaries and branches into focus.

Human engineering is, however, recognizable as a specialized, scientific enterprise if one looks to the turn of this century when certain psychologists began to apply some of the findings of their psychological experiments to work situations. It is interesting to notice that the main research targets at that time were work conditions in various industries and the investigation of human potentialities and limitations; the study of skills, fatigue and so on. All sorts of data were gathered on the effects of heat, light, noise, etc., on performance at work and in the case of human skills the chief interests lay in the study of individual tasks. The pioneer work of Taylor and Gilbreth in the field of motion and time study helped to draw attention to the value of the analysis of jobs by skilled technicians. Even now the pattern of problems may be startlingly similar when, for example, attempts are made to determine the optimal situation or dimension for dials, pointers, conveyors, starting buttons, handles etc. The current aim is to eliminate or at least reduce machine errors, man errors and man-machine errors and to construct a system which will best fit the product and the characteristics of the individual workers turning it out. Such goals, if achieved, are likely to save considerable time and expense in personnel selection and training.

Of course special complications arise with the introduction of automatic processes and with increases in mechanization, but it is unlikely that such industrial changes will be applicable to all types of work. It is probable that many jobs will continue to require either continuous or intense muscular effort, and in such cases more detailed study of work methods is likely to pay rich rewards in efficiency and conservation of effort. Moreover, it may not be desirable or practicable to automate jobs involving a limited number of products of any one kind, and maintenance work or individual assembly are unlikely to alter radically in the near future.

It is probably no exaggeration to say that the urgent need to solve certain wartime problems some twenty years ago really put human engineering on its feet. In addition to medical research related to the effects of excessive heat and cold, to fatigue and stress in aircraft pilots, to clothing design for extreme climatic conditions, the general increase in the complexity of wartime military equipment raised acute problems directly related to safety, fighting efficiency and to survival itself. Almost parallel developments took place in this country and in the United States of America—the designers of equipment being charged with the task of allowing for human characteristics in their plans. Multiple teams of scientists were set up and many of their findings were carried over into peace-time situations. During the last ten years or so the field of ergonomics has blossomed again and shows signs of being as productive in its industrial roles in peace as it was in relation to the Armed Forces in war. Yet even at the present time the majority of psychologists employed on problems of human engineering in this country are to be found in the Government Service. It is true, however, that some of the research carried out elsewhere makes use of experimental findings which are, at least in part, based on this approach. Various industrial organizations have set up separate departments to investigate their own particular problems.

3. Research Methods and Techniques

It should now be apparent that although human engineering may be regarded as a definitive field with some uniformity of purpose, it would nevertheless be inaccurate to describe the nature of its methods and techniques as anything other than varied. Both the nature of the problems themselves and the fact that contributions come from scientists in different disciplines are largely responsible for this high variability. Each scientist tends to bring the techniques and expertise of his own profession to the situation; and to express the problem and its solution, if not in a private "jargon", then frequently at least, in terms of concepts from his own field of knowledge. Such problems of communication are not exclusive to human engineering but must not be overlooked for all that.

Yet, at base, there is indeed some common foundation in methodology if not technique to which all the various specialists would aspire, if not attain. This can adequately and acceptably be described as the systematic approach of the scientist to the problem and its solution. Although it would be inappropriate here to outline the nature of scientific methods, it is vital to remember that it is as important to ask the "right" questions as it is to apply scientifically rigorous methods for their unequivocal answer. The isolation of a problem area, for example, is typically only a primitive stage in a complex process. A problem area may eventually appear to be made up of quite dissimilar problems meriting research using very different methods and techniques. Many problems are pseudo-problems; others cannot be handled by the techniques available at the time and have to be rephrased or otherwise refined. It is imperative that in the early phases of exploring likely hypotheses or following-up "hunches", the research worker looks beyond the immediate situation to assess whether he has the necessary verbal, experimental and mathematical tools for handling the problem from conception to solution. Too much research activity is wasted because of this form of shortsightedness. Too many data, respectably amassed, lie useless for want of frameworks into which they can fit.

It is not enough then, to be systematic and rigorous about observation and measurement; not enough to vary only experimental conditions and control extraneous factors; not enough to practise precise standards of recording data. Necessary as they certainly are, they may be insufficient to solve the problem, support or reject a set of hypotheses, or enable predictions to be made. In scientific research the crucial steps are taken long before the experimental apparatus is set up and long after it has been dismantled.

3.1. *The "Human Factor"*

The notion of the "human element" or "human factor" as a source of error or as a causative factor in a situation is often expressed in everyday speech with a sort of desperation and pessimism implying the futility of any remedial action as a consequence of the unpredictability of human behaviour. At other times its utterance is at least influenced by the vagueness and irregularity of human attributes and patterns of response. The student of human perception, learning, remembering, thinking, emotion etc. concedes the lack of precision in his predictions in contrast to the physicist's pronouncements about, say, the behaviour of bodies falling in a vacuum, but, in addition to re-examining his theories and concepts and improving the accuracy and resolving power of his techniques, he could, with some justification, ask that allowances be made for the greater complexity of the central object under study and for his attempt to maintain some degree of reality in his investigations.

Taking the second point first, we should notice that different branches of scientific enquiry must learn to steer their different courses between the precise, quantitative and objective at one extreme, and the less rigorous, descriptive study of "real life" at the other. The former can be highly artificial and ultra-specific; the latter, even anecdotal. Certainly, these are perhaps exaggerated extremes and most scientific investigations tend to lie somewhere in between, but it should be noted that only the most formal of scientists can exclude all but the former and that exclusive and consistent use of the latter tends usually to take us outside the boundaries of science altogether.

In the case of human engineering, theoretical exposition sometimes apart, the approach to the solution of a problem is frequently dictated by the nature of that problem. For example, in some situations the effect of a particular layout of instruments or arrangement of controls on the performance of a particular task may be the target for study. Or it may be that the research is concerned with tracing the pattern of development of certain skills when men operate a particular machine. In such cases the most suitable type of study is likely to consist of a series of laboratory experiments, perhaps carried out in some test room constructed to rigid specifications. Carefully selected human subjects perform under different but controlled conditions and results are treated to suitable statistical investigation. Such procedures, if properly carried out, have the accuracy and precision of scientific experimentation but, typically, neither take account of long-term effects nor are directly applicable to the shop floor where work conditions, the nature of supervision and factors of job experience etc. are liable, directly or indirectly, to influence performance on the task.

Solutions to other kinds of problems, however, have been found to be more suitably obtained by approaches which, characteristically, have less precision but which have the advantage of a closer approximation to reality. This kind of approach is seen in the situation in which the problem is investigated in the natural setting of, say, the factory floor. Here, the crucial features of the problem are seen to involve the effects of various factors over an extended period of time, and also, perhaps, where the influence of changes in work method or work conditions is to be explicitly related to psychological phenomena. The early studies of work environment measuring the subjective influences of different decorative schemes and levels of illumination in workshops are typical examples of this approach.

Returning now to the question of the complexity of the human subject in his role as worker or operator, we should be aware of the difficulties of prediction not only because of differences between the responses of different individuals in an identical situation, but because of the response differences of the same individual at different times. Although certain techniques and methods can increase the efficiency of prediction under variable circumstances, inferences and statements of probability have often to be made at a low level of confidence. Yet one must not conclude pessimistically that all human behaviour is infinitely variable. Generalizations are not impossible. In some spheres of human activity the margin of predictive error is greater than in others—perhaps nowhere greater than in situations where factors of individual motive, attitudes etc. are crucial. Even in the "dummy" experimental set-up we cannot afford to ignore such characteristics. In the natural setting of the workshop, although we may be dealing with actual workers with their own characteristic personalities, we have to take account of motivational changes occurring not only as a consequence of the manipulation of factors in the investigation, but pay attention to alterations in conditions extraneous to that investigation. There is ample evidence, for example, that the mere presence of an investigator can alter patterns of motivation and hence the performance of those who are under scrutiny.

There are, indeed, many other problems specific to the measurement of human attributes in both real and artificial settings, but these cannot be dealt with at this time. It is perhaps sufficient, here, to point to the limitations such problems impose upon methods and techniques which, in themselves, may be far from perfect. In practice, of course, methods and techniques are frequently modified to take account of new information, or are replaced by more accurate or more appropriate successors.

4. Research Areas in Human Engineering

The field of human engineering is vast, diverse and continually expanding so that it is quite impracticable to do much more than sample some few areas of investigation in an article of this kind. It would be incorrect to assume that the areas selected for attention were necessarily more important or more representative than other areas not so selected. For a broader and more detailed coverage of the field a number of excellent texts are available, e.g. Chapanis[1] and McCormick.[2]

4.1. *Vision, Illumination and Colour*

The frequency of studies in the visual field as applied to work situations rests, obviously, in the fact that visual processes play a major role in many jobs. Also, the range of visual tasks is very wide indeed—some close, precise and fatiguing; others coarse but nevertheless vital. One cannot ignore the visual demands of occupations although there is clearly a wide gulf between the visual skills of the railway engine driver or the jeweller and those of the steelman or builder's labourer. In many cases visual demands may involve measurements of acuity, the perception of depth, adaptation to darkness, convergence, and so on.

The closely allied field of illumination is clearly one of the critical environmental factors related both to efficiency and to satisfaction in work. Central in any investigation of suitable illumination is the concept of the critical level for any particular task. It is worthless merely to have opinions that the workers are or are not seeing "well". Rather, one must measure the amount of illumination in relation to this critical level—the amount of illumination beyond which no further increase in productive efficiency is obtained. Standards by which the adequacy of illumination level are assessed tend to vary with the working environment, but many of the relevant factors are known. These, of course, include personal preferences as well as factors which are more easily quantified. Illumination level is not only associated with output but with factors which tend to have negative effects, e.g. accident rate. Too high a level of illumination can result in glare, which itself can lower efficiency and produce visual discomfort.

Apart from the necessity of assessing the significance of certain standards of colour vision for adequate performance on the job, the use of colour coding for controls etc., provides a valuable safety measure. Colour coding provides a means of increasing the accuracy of identification as well as of discrimination. In general, the attention-getting characteristics of colour and their provision of visual contrast are attributes that can be harnessed to effect when designs for equipment are being considered. There is little real evidence that colours directly influence the emotional behaviour of individuals—in general, their effects have tended to be overemphasized. This does not mean, of course, that there are no such things as colour preferences, very variable as they may be. These preferences seem to depend upon the nature of the coloured object and its functional significance. Nevertheless, there appears to be a more stable set of preferences for any one individual in a variety of situations, than there is a stable preference among a variety of individuals in one specific setting.

4.2. *Auditory Communication and Noise*

A vast literature on auditory mechanisms and processes is available. In many respects the kind of auditory phenomena meriting investigation are similar to those in the visual field—the auditory demands of jobs, the characteristics of hearing structures and processes in man and their operation in different work situations etc. Just as visual communication is important in some jobs, so is auditory communication important in others. Much research has been carried out on the nature of messages, the medium in which they are transmitted and on the receiver himself. It is not possible to summarize the results of such research here but it should be obvious that this is likely to become a field of even greater significance as auditory displays multiply.

Research on noise and the interpretation of data therefrom have similarly been extensive, but progress in regard to effects, physiological and psychological, has been slow. The relation between noise and work performance is more controversial than many would suppose and it is probably not being over-cautious to state that little definitive information is available about the detrimental effects of noise except at high intensities. Even although the effects of noise of high intensity, high frequency or where there are unpleasant reverberations, do not attain a level which is physiologically harmful, there always remains for consideration effects which are "psychologically" harmful to worker effort. It is precisely in the latter case where the evidence is most inconclusive. Whenever a practical situation arises in which noise level is likely to be contributory to a fall in output or, say, morale, steps should be taken to measure its level, its likely effects and to devise appropriate controlling devices.

4.3. *Atmospheric Conditions*

Many of the pioneer studies by psychologists in industry were concerned with the nature of the work environment, emphasis often being placed on the physical work conditions. Studies of atmospheric conditions come under this heading and usually deal either with the effects of different concentrations of substances on performance (e.g. lack of oxygen, level of impure gases etc.), or with the effects of different levels of humidity, temperature etc. Clearly, these effects are measurable at a physiological level—what, for example, are individual tolerances to certain atmospheric factors, or how do different concentrations of substances in the air affect basic metabolic processes? But they are also measurable at a psychological level; in terms of performance or level of skill, for example. A lot of information is available about the effects of atmospheric extremes; these data provide useful, basic knowledge for the physiologist as well as for the Polar explorer. There seems to be a great need, however, for more attention to be paid to the effects of relatively minor discomforts in the work situation. Many of our heavy industries should pay attention to the effect of various atmospheric "stresses" on human performance, e.g. heat stress in foundries.[3]

4.4. *Design of Displays, Controls and Equipment*

Displays are devices whereby information is provided through the medium of symbols. A speedometer, a clinical thermometer or a clock are all examples of visual displays transmitting information indicating a certain state of affairs. Technically, displays need not be visual. Whether or not the information transmitted is critical (and it frequently is), it seems only good sense to arrange displays in such a way that the information can be read off easily and accurately. Since it has frequently been demonstrated that the main source of error in performance can be traced directly to misreading of instruments, it is worthwhile to investigate the characteristics of dials and instruments so that their design eliminates or reduces the error factor. The present increase in development of automatic processes is often associated with an increase in the number of displays, making it all the more necessary to study their design. Sleight,[4] for example, studied the legibility of five different types of dials; other research has concentrated upon such aspects of display as the choice of dial units and their markings, types of pointers, number and spacing of markers, choice of numerals and letters etc. One must add as a caution that what appears to be a good design in one situation may be a poor design in another.

The same kind of considerations that apply to displays equally apply to controls. Just as we must take account of arrangement, legibility etc. of displays, we must devise the most appropriate control mechanism for a specific situation. Some years ago, Jenkins[5] studied the shape coding of lever knobs in order to find out which shape of knob was most easily identified by touch without confusion with others. In other situations visual coding of controls may be necessary. One could easily find many other examples of research into the design of controls; what is perhaps more important is to point out that such considerations are often neglected.

The design of equipment is, of course, related to the design of displays and the arrangement of controls, and, like these activities, often is arrived at using criteria and principles which ensure that the equipment is mechanically sound and even aesthetically desirable. Functionally, however, the equipment may leave much to be desired. Not only must the nature of the equipment be considered; so also must be its arrangement. The optimal arrangement of equipment should have the comfort and convenience of the operators as a goal not inferior to that of greater efficiency at work. The truth is that the latter depends to a large extent upon the former.

There are encouraging signs that industry itself is paying more attention to such factors. Singleton and Simister,[6] for example, studied the design and layout of machinery in the shoe manufacturing industry. Part of their research showed that the addition of extra foot controls on a sewing machine enabled the operator to keep the needle in the "up" position without removing the right hand from the work. This modification increased productivity quite markedly and resulted, among other things, in better control of the machine.

It would be a mistake to assume that research into the design of equipment is restricted to manufacturing industries. Scientific applications of human engineering can be found in a wide variety of situations. Experimental studies on the design of motor vehicles, for example, have dealt with such different problems as the design of headlights and the important safety factor of visibility from the driving seat. Again, in the field of electronic equipment, studies of fault-finding have been carried out. Dale,[7] for example, analysed fault-finding behaviour experimentally in terms of the various strategies used and their relative effectiveness, the factor of method of presentation and, of course, individual differences.

The awareness of the need for various specialists to take account of the field of human engineering can be seen in the article by Shackel.[8]

4.5. *Motor Performance*

Sooner or later in most problems in the field of human engineering we are faced with an analysis of the performance of an individual or a group of individuals on some job or another. Sooner or later questions as to accuracy, speed, strength etc. of movement will arise. At base, these are physiological questions for which many adequate answers are available. Sometimes, however, we need not resort to this level of analysis as long as we can accept a classification of performance which is discriminative; for example, repetitive as contrasted with serial movement. There are many ways in which we can classify movements and many ways in which their analysis is helpful in, say, problems of equipment design, control design etc. In general, the performance of motor activities is influenced by such factors as angles, speeds, directions, positions, and so on. These factors are usually assessed by laboratory experiment, for only here are we able to achieve precision of control. The nature of a skilled performance is likely to defy casual observation. The latter has too often been the only method of analysis.

5. Conclusion

It is always a difficult task to stand aside and get an overall view of a field that is continually growing. The field of human engineering is somewhat like this in that it frequently slips out of focus. Yet it can be said that it is a lively and expanding field, and that many different practitioners operate within it. It is a field which owes much to the pioneers of research in skill, fatigue and work conditions, but now includes many other topics as well. Only a small sample of problem areas was considered here. Perhaps the areas which are expanding most at the present time are those related to the design of complex equipment and its control. Not only complex equipment receives research attention however. Designing rockets and radar warning systems and launching men into space may well be crucial activities both now and in the future, but there are more down-to-earth applications too. Since they range from the study of the effects of high temperature conditions in foundries and the design of protective clothing and safety devices to the more placid office, school and home environments, one cannot complain that human engineering only benefits some humans.

Although the emphasis has here been placed on the practical application of results from studies in human engineering, this should not be taken as an implication that the data preclude theoretical treatment, or even discourage the advancement of theory. Despite the fact that at the present time there are more data than explanations and generalizations about them, theoretical advancement is already becoming evident—perhaps especially because of the contribution of theoretical models constructed from concepts in information theory and in engineering. Perhaps the union between concepts from studies in human behaviour and those from the physical sciences will be as productive to theory as it has been to practice, and it may well be that the use of a common language by applied psychologists and by engineers and physical scientists will encourage co-operation and facilitate communication between them. These alone would be worthy achievements in an area where the disciplines are so closely related.

6. References

1. A. Chapanis, "Research Techniques in Human Engineering". (The Johns Hopkins Press, Baltimore, 1959.)
2. E. J. McCormick, "Human Engineering". (McGraw-Hill, New York, 1957.)
3. D. Turner, "Heat stress in non-ferrous foundries", *British Journal of Industrial Medicine*, **15**, pp. 38–40, 1958.
4. R. B. Sleight, "The effect of instrument dial shape on legibility", *Journal of Applied Psychology*, **32**, pp. 170–88, 1948.
5. W. O. Jenkins, "The tactual discrimination of shapes for coding aircraft-type controls". In P. M. Fitts (Ed.) "Psychological Research on Equipment Design", pp. 199–205. (U.S. Government Printing Office, Washington, 1947.)
6. W. T. Singleton and R. Simister, "The design and layout of machinery for industrial operatives", *Occupational Psychology*, **4**, pp. 234–242, 1957.
7. H. C. A. Dale, "Fault-finding in electronic equipment", *Ergonomics*, **1**, pp. 356–385, 1958; "Priorities in the training of maintenance technicians", *J. Brit.I.R.E.*, **20**, pp. 715–7, September 1960.
8. B. Shackel, "Human engineering and electronics", *British Communication and Electronics*, **4**, pp. 350–356, 1957.

Manuscript received by the Institution on 10th November 1961 (Paper No. 738).

IMPROVEMENTS IN BROADCASTING COVERAGE IN GREAT BRITAIN

V.H.F. Sound Transmissions from Station near Oxford

The v.h.f. sound transmitters at the B.B.C.'s new transmitting station at Beckley, near Oxford, were brought into service on Monday 28th May. This station is one of several combined television and v.h.f. sound broadcasting stations which are being built to extend and improve the coverage of the B.B.C.'s services. The television transmitter is already in operation.

As the service area of the v.h.f. sound transmitters includes parts of the B.B.C.'s Midland and West Regions, the Beckley station transmits both the Midland and West of England Home Services in addition to the Light Programme, the Third Programme and Network Three. Horizontal polarization is used.

The Beckley station, which is designed to work unattended, will receive its sound programmes by radio, the West of England Home Service from Wenvoe or Rowridge, and the other three services from Sutton Coldfield. It will extend the B.B.C.'s v.h.f. sound service to a quarter of a million people and provide improved reception for a further quarter of a million people in an area which includes Oxford, Bicester, Witney, Swindon, Wantage and Aylesbury.

New Band III Aerial for London

The Independent Television Authority's London station is to have a new transmitting aerial with the object of increasing the strength of signals received in north and north-west London, Middlesex, Hertfordshire and Buckinghamshire. The existing installation at Croydon will be replaced by a 500 ft tower, carrying initially one aerial array, of 80-ft aperture, to transmit in Band III the existing programmes on Channel 9. The array will use vertical polarization and will be of similar basic type to the arrays at I.T.A. transmitters at Dover, Caradon Hill, Stockland Hill and Lichfield. In addition, the tower will have provision for the fitting of a second Band III 80-ft aperture array, and will be capable of supporting two u.h.f. aerials for transmitting in Bands IV and V.

Construction of the mast has already advanced to a stage at which the structure itself is causing some limited interference to the radiation from the present aerial. This leads to ghosting and attenuation roughly along a line from Croydon to Southend which will continue, possibly varying in direction and intensity, until the new aerial comes into service in the autumn.

The array is being built by E.M.I. Electronics, and the tower by British Insulated Callender's Cables.

NORTH-EASTERN SECTION REPORT

The North-East of England is becoming increasingly radio and electronics minded. The traditional industries of coal mining, shipbuilding and heavy engineering are making more and more use of electronics to aid them in production while engineering, radio and optical works within the area produce equipment which is well known throughout the world. One of the largest computer installations in the British Isles is in the North-East, and the University and Technical Colleges have more students studying radio and electronics than ever before.

It is therefore not surprising that membership of the North-Eastern Section has risen during the past year, and that the meetings held at the Institute of Mining and Mechanical Engineers have been well attended.

The 1961–62 session commenced with a paper by Mr. A. J. Rees—"V.H.F. Communications Receivers and Transmitters using Transistors". This lecture was very well received and subsequently given before the North-Western Section on the 2nd November, 1961. It was reviewed in the *Journal* in February 1962.

A very interesting paper was read by Mr. G. D. Browne on the 8th November, entitled "A Pulse Time Multiplex System for Stereophonic Broadcasting". His demonstration of the "sampling" of left and right audio channels over one v.h.f. carrier was very impressive. The paper was published in the *Journal* in February 1962.

At the December meeting the audience heard Dr. T. H. Wilmshurst of the Department of Electronics, Southampton University, talk about "Masers and Parametric Amplifiers". The lecturer outlined the principles involved and some of the applications to which they are put. "Lasers" were also dealt with during the discussion that followed.

The title of the lecture at the first meeting of the new year was "Industrial Electronic Temperature Measurement and Control". This was delivered by Mr. B. N. Ellis, a graduate member of the Section, who introduced his lecture by a close examination of suitable sensing elements which can convert the heat available in an industrial process into a suitable electrical signal to operate different forms of electronic instruments. Various systems of control techniques were then presented. The paper concluded with a look at the developments that are likely in the immediate future in this field of technology and with a discussion of some of the ergonomic aspects.

Dr. L. Molyneux of the Physics Department, King's College, University of Durham, presented the February paper on "Some Physical and Physiological Signal-to-Noise Ratios". He gave a number of examples where a high signal-to-noise ratio is very important which included medical applications, the detection of earth tremors and testing the magnetism of rocks. Details were given of the special equipment necessary for these measurements.

The March lecture was delivered by Dr. E. A. Freeman and was entitled "Performance and Design Considerations of High Gain D.C. Amplifiers to be used in Analogue Computers". After showing how an ideal infinite gain amplifier may be used as a computing element and illustrating the various forms of calculation that could be performed, practical limitations were introduced and their effects on accuracy discussed. Methods of stabilizing a practical amplifier and the effects of finite gain, noise, drift and grid current were also described.

An extension to the usual activities of the Section occurred on the 10th April when a meeting was held in conjunction with The School of Signals at Catterick Camp. Mr. J. B. Wilson, M.Sc., delivered a paper on "Jodrell Bank" in which he described some of the ways in which radio astronomy has extended our knowledge of the universe. An appreciative audience of about a hundred students and staff of the School included the Commandant, Brigadier H. F. McGill, and senior members of his staff, and the Captain of the Royal Navy School of Signals (H.M.S. *Mercury*), Captain D. E. Bromley Martin. The meeting was arranged with the close co-operation of Major N. McIntyre (Associate Member) and in view of its success, it is intended to hold further meetings at Catterick in the coming session. This year's Section visit was in fact to the School of Signals and took place on 16th June. Members were able to examine a wide range of communication equipment.

A paper entitled "The Design of High Quality Sound Reproducing Equipment", followed the Annual General Meeting of the Section which took place on 11th April. Mr. K. Davin and Mr. F. C. Gibson showed the progress of sound reproducing equipment from the early days of disc recording and sound broadcasting. Basic requirements of signal sources, amplifiers and loudspeakers and enclosures were discussed together with detailed accounts of present-day techniques. Demonstrations illustrated the various aspects covered from the early moving iron loudspeakers to modern equipment. The audience was also entertained by a very good stereophonic music demonstration.

In making its plans for the coming session the North-Eastern Section Committee wishes to make it more widely known that there is encouragement for Brit.I.R.E. members to read papers before the Section; the Hugh Brennan Premium is awarded for the best paper presented by a North-Eastern Section member and published in the *Journal*. J. W. O.

U.D.C. 621.373.431

A Two-State Device with Two Inductively Coupled Colpitts Oscillators

By

B. R. NAG,

M.Sc.(Tech.), M.S., D.Phil. †

Summary: Characteristics of a system consisting of two inductively-coupled Colpitts oscillators are studied. It is shown that the system may be used as a two-state device and gives a switching time better than conventional circuits.

1. Introduction

The suitability of oscillators having more than one mode of oscillation for use as storing devices has been examined by Edson,[1] Disman[2] and others. An oscillator with two tuned circuits was first studied by van der Pol[3] and subsequently by Fontana,[4] Skinner,[5] Schaffner[6] and other workers. The present author examined the characteristics of a system consisting of two tuned anode oscillators coupled by a common capacitance and found that the system has two stable states and may be switched over from one state to the other by pulses. It may, therefore, be used as a two-state device. The switching time of the device is determined by the frequency of oscillation. However, as the frequency of oscillation is increased with a view to reduce the switching time, the tuned circuit parameters eventually become comparable to the interelectrode capacitances. The oscillators can no longer be considered to be of the tuned anode type. In effect, the oscillators assume the form of tuned-grid, tuned-anode oscillators with mutual coupling between the anode and grid coils. For this reason, it is rather difficult to operate the system at the highest oscillation frequency permitted by the tube inter-electrode capacitances, and also the analysis of the system is quite complex.

It is known that a comparatively higher oscillation frequency may be realized by operating the oscillators in the Colpitts form. This fact prompted the author to examine the possibility of constructing a two-state device by two inductively-coupled Colpitts oscillators, with a much smaller switching time than other devices. The oscillator system is then described by two non-linear equations each of the third order. Such a system is studied in the present paper. The describing differential equations are deduced in Section 2.1. Solutions to the differential equations are worked out in Section 2.2. Suitable values of circuit parameters for operating the system as a two-state device are obtained in Section 2.3. Experimental results confirming the analytical conclusions are presented in Section 3.

2. The Oscillator System and its Analysis

2.1. *Describing Differential Equations of the Oscillator System*

The equivalent circuit of the oscillator system considered is shown in Fig. 1. The different symbols used in the following analysis are defined in the figure. Considering this equivalent circuit one may write the following equations relating the voltages and currents in the different meshes:

$$V_{11} = V_1 + L_1 C_{11}\frac{d^2V_1}{dt^2} + MC_{22}\frac{d^2V_2}{dt^2} \qquad \ldots(1)$$

$$g_{m1}V_1 = \frac{V_{11}}{R_1} + C_1\frac{dV_{11}}{dt} + C_{11}\frac{dV_1}{dt} \qquad \ldots\ldots(2)$$

$$V_{22} = V_2 + L_2 C_{22}\frac{d^2V_2}{dt^2} + MC_{11}\frac{d^2V_1}{dt^2} \qquad \ldots(3)$$

$$g_{m2}V_2 = \frac{V_{22}}{R_2} + C_2\frac{dV_{22}}{dt} + C_{22}\frac{dV_2}{dt} \qquad \ldots\ldots(4)$$

It is assumed that no current flows from grid to cathode in each valve.

Combining (1) with (2) and (3) with (4) and after simplification one obtains

$$\frac{d^3V_1}{dt^3} + \frac{1}{C_1R_1}\frac{d^2V_1}{dt^2} + \frac{1}{L_1}\left(\frac{1}{C_1}+\frac{1}{C_{11}}\right)\frac{dV_1}{dt} + \left(g_{m1}+\frac{1}{R_1}\right)\frac{1}{L_1C_1C_{11}}V_1 + \frac{MC_{22}}{L_1C_{11}}\frac{d^3V_2}{dt^3} + \frac{MC_{22}}{L_1C_{11}C_1R_1}\frac{d^2V_2}{dt^2} = 0 \qquad \ldots\ldots(5)$$

$$\frac{d^3V_2}{dt^3} + \frac{1}{C_2R_2}\frac{d^2V_2}{dt^2} + \frac{1}{L_2}\left(\frac{1}{C_2}+\frac{1}{C_{22}}\right)\frac{dV_2}{dt} + \left(g_{m2}+\frac{1}{R_2}\right)\frac{1}{L_2C_2C_{22}}\,{}_2 + \frac{MC_{11}}{L_2C_{22}}\frac{d^3V_1}{dt^3} + \frac{MC_{11}}{L_2C_{22}C_2R_2}\frac{d^2V_1}{dt^2} = 0 \qquad \ldots\ldots(6)$$

† Institute of Radio Physics and Electronics, University of Calcutta.

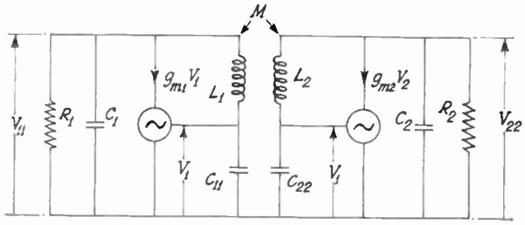


Fig. 1. Equivalent circuit of the oscillator system.

It is assumed that the mutual conductances of the valves are functions of the grid voltage only and may be written as

$$g_{m1} = a_{01}+a_{21}V_1^2 \qquad \ldots\ldots(7)$$

$$g_{m2} = a_{02}+a_{22}V_2^2 \qquad \ldots\ldots(8)$$

It may be remarked that the above assumption is not fully justified. However, if the mutual conductances are measured under the dynamic condition by putting a resistance in the anode circuit simulating the tuned circuit impedance and considering that other terms in the polynomial expansions for g_{m1} and g_{m2} are relatively unimportant in the limiting characteristics of the oscillators, eqns (7) and (8) may be taken to be reasonable assumptions. Substituting (7) and (8) in (5) and (6) respectively and putting

$$\frac{1}{C_1R_1} = a_1, \quad \frac{1}{L_1}\left(\frac{1}{C_1}+\frac{1}{C_{11}}\right) = \omega_1^2, \quad \frac{MC_{22}}{L_1C_{11}} = k_2$$

$$\frac{1}{C_2R_2} = a_2, \quad \frac{1}{L_2}\left(\frac{1}{C_2}+\frac{1}{C_{22}}\right) = \omega_2^2, \quad \frac{MC_{11}}{L_2C_{22}} = k_1$$

one gets

$$\frac{d^3V_1}{dt^3}+a_1\frac{d^2V_1}{dt^2}+\omega_1^2\frac{dV_1}{dt}+\left(a_{01}+\frac{1}{R_1}+a_{21}V_1^2\right)\frac{1}{L_1C_1C_{11}}V_1+k_2\frac{d^3V_2}{dt^3}+a_1k_2\frac{d^2V_2}{dt^2} = 0 \qquad \ldots\ldots(9)$$

$$\frac{d^3V_2}{dt^3}+a_2\frac{d^2V_2}{dt^2}+\omega_2^2\frac{dV_2}{dt}+\left(a_{02}+\frac{1}{R_2}+a_{22}V_2^2\right)\frac{1}{L_2C_2C_{22}}V_2+k_1\frac{d^3V_1}{dt^3}+a_2k_1\frac{d^2V_1}{dt^2} = 0 \qquad \ldots\ldots(10)$$

In order that the system should oscillate it is necessary that

$$\left(a_{01}+\frac{1}{R_1}\right)\frac{1}{L_1C_1C_{11}} \geqslant a_1\omega_1^2 \qquad \ldots\ldots(11)$$

and
$$\left(a_{02}+\frac{1}{R_2}\right)\frac{1}{L_2C_2C_{22}} \geqslant a_2\omega_2^2 \qquad \ldots\ldots(12)$$

In practice, the oscillators will be so arranged to satisfy the above inequalities that one may write

$$\left(a_{01}+\frac{1}{R_1}\right)\frac{1}{L_1C_1C_{11}} = a_1\omega_1^2+b_1 \qquad \ldots\ldots(13)$$

$$\left(a_{02}+\frac{1}{R_2}\right)\frac{1}{L_2C_2C_{22}} = a_2\omega_2^2+b_2 \qquad \ldots\ldots(14)$$

In (13) and (14) b_1 and b_2 will be positive and, in general, small.

Substituting (13) and (14) in (9) and (10) and putting

$$\frac{a_{21}}{L_1C_1C_{11}} = -c_1, \qquad \frac{a_{22}}{L_2C_2C_{22}} = -c_2$$

one gets

$$\frac{d^3V_1}{dt^3}+a_1\frac{d^2V_1}{dt^2}+\omega_1^2\frac{dV_1}{dt}+a_1\omega_1^2V_1+k_2\frac{d^3V_2}{dt^3}+ + k_2a_1\frac{d^2V_2}{dt^2} = -b_1V_1+c_1V_1^3 \qquad \ldots\ldots(15)$$

$$\frac{d^3V_2}{dt^3}+a_2\frac{d^2V_2}{dt^2}+\omega_2^2\frac{dV_2}{dt}+a_2\omega_2^2V_2+k_1\frac{d^3V_1}{dt^3}+ + k_1a_2\frac{d^2V_1}{dt^2} = -b_1V_2+c_2V_2^3 \qquad \ldots\ldots(16)$$

It may be noted that a_{21} and a_{22} are usually negative, so that c_1 and c_2 are positive. Equations (15) and (16) are the describing differential equations of the system.

2.2. *Solution of the Describing Differential Equations*

The solution of the equations is worked out following the van der Pol's method of approximation. It is assumed that the solutions for V_1 and V_2 are given by

$$V_1 = A_1\sin(\omega_{10}t+\phi_1)+B_1\sin(\omega_{20}t+\theta_1) \ldots\ldots(17)$$

$$V_2 = A_2\sin(\omega_{10}t+\phi_2)+B_2\sin(\omega_{20}t+\theta_2) \ldots\ldots(18)$$

In eqns (17) and (18), A_1, A_2, B_1, B_2, ϕ_1, ϕ_2, θ_1 and θ_2 are, in general, functions of time. The frequencies ω_{10} and ω_{20} are the frequencies of free oscillation when the right-hand side terms of (15) and (16) are absent and are given by

$$\omega_{10}^2 = [\omega_1^2+\omega_2^2- -\{(\omega_1^2-\omega_2^2)^2+4k_1k_2\omega_1^2\omega_2^2\}^{\frac{1}{2}}]/2(1-k_1k_2) \ldots(19)$$

$$\omega_{20}^2 = [\omega_1^2+\omega_2^2+ +\{(\omega_1^2-\omega_2^2)^2+4k_1k_2\omega_1^2\omega_2^2\}^{\frac{1}{2}}]/2(1-k_1k_2) \ldots(20)$$

Substituting (17) and (18) in (15) and (16), neglecting the derivatives of the amplitudes and phases of orders higher than one, the products of the derivatives and the coefficients of the right-hand side terms in (15)–(16) and the harmonics generated by non-linear terms, and collecting the sine and cosine terms one obtains (Appendix 1) for the components of frequency ω_{10},

$$n\omega_{10}(a_1+a_2)\frac{dA_1}{dt}-n(\omega_{10}^2-a_1a_2)A_1\frac{d\phi_1}{dt}=a_2(\omega_2^2-\omega_{10}^2)\mu_1A_1+k_2\omega_{10}^2\mu_2A_2(\omega_{10}\sin\phi-a_1\cos\phi) \quad \ldots(21)$$

$$n(\omega_{10}^2-a_1a_2)\frac{dA_1}{dt}+n\omega_{10}(a_1+a_2)A_1\frac{d\phi_1}{dt}=\omega_{10}(\omega_2^2-\omega_{10}^2)\mu_1A_1-k_2\omega_{10}^2\mu_2A_2(\omega_{10}\cos\phi+a_1\sin\phi) \quad \ldots(22)$$

$$n\omega_{10}(a_1+a_2)\frac{dA_2}{dt}-n(\omega_{10}^2-a_1a_2)A_2\frac{d\phi_2}{dt}=a_1(\omega_1^2-\omega_{10}^2)\mu_2A_2-k_1\omega_{10}^2\mu_1A_1(\omega_{10}\sin\phi+a_2\cos\phi) \quad \ldots(23)$$

$$n(\omega_{10}^2-a_1a_2)\frac{dA_2}{dt}+n\omega_{10}(a_1+a_2)A_2\frac{d\phi_2}{dt}=\omega_{10}(\omega_1^2-\omega_{10}^2)\mu_2A_2-k_1\omega_{10}^2\mu_1A_1(\omega_{10}\cos\phi-a_2\sin\phi) \quad \ldots(24)$$

where

$$n=2\omega_{10}(1-k_1k_2)(\omega_{10}^2-\omega_{20}^2)$$

$$\mu_1=\tfrac{3}{4}c_1(A_1^2+2B_1^2)-b_1, \qquad \mu_2=\tfrac{3}{4}c_2(A_2^2+2B_2^2)-b_2, \qquad \phi=\phi_2-\phi_1$$

It may be noted that

$$\frac{1}{A_1}\frac{dA_1}{dt}=\frac{1}{A_2}\frac{dA_2}{dt}, \qquad \frac{d\phi_1}{dt}=\frac{d\phi_2}{dt}$$

Eliminating $\frac{dA_1}{dt}, \frac{dA_2}{dt}, \frac{d\phi_1}{dt}, \frac{d\phi_2}{dt}$ one obtains

$$\sin\phi=\frac{(a_1-a_2)\mu_1\mu_2A_1A_2[k_1A_1^2(\omega_1^2-\omega_{10}^2)-k_2A_2^2(\omega_2^2-\omega_{10}^2)]}{\omega_{10}[\mu_2^2k_2^2A_2^4(\omega_{10}^2+a_1^2)-\mu_1^2k_1^2A_1^4(\omega_{10}^2+a_2^2)]} \quad \ldots\ldots(25)$$

$$\begin{aligned}\cos\phi=&[(\omega_{10}^2+a_1a_2)\{k_1A_1^2(\omega_1^2-\omega_{10}^2)-k_2A_2^2(\omega_2^2-\omega_{10}^2)\}\mu_1\mu_2A_1A_2+(\omega_{10}^2+a_1^2)\mu_2^2k_2A_2^3A_1(\omega_1^2-\omega_{10}^2)-\\&-(\omega_{10}^2+a_2^2)\mu_1^2k_1A_1^3A_2(\omega_2^2-\omega_{10}^2)]\times[\omega_{10}^2\{\mu_2^2k_2^2A_2^4(\omega_{10}^2+a_1^2)-\mu_1^2k_1^2A_1^4(\omega_{10}^2+a_2^2)]^{-1} \quad \ldots\ldots(26)\end{aligned}$$

On combining (25) and (26)

$$\begin{aligned}&[k_1A_1^2(\omega_1^2-\omega_{10}^2)-k_2A_2^2(\omega_2^2-\omega_{10}^2)]\times\\&\times\left[\mu_2^4\frac{k_2^2}{k_1}A_2^6(\omega_{10}^2+a_1^2)(\omega_1^2-\omega_{10}^2)+\mu_1^4\frac{k_1^2}{k_2}A_1^6(\omega_{10}^2+a_2^2)(\omega_2^2-\omega_{10}^2)+\mu_1^2\mu_2^2(\omega_{10}^2+a_1^2)(\omega_{10}^2+a_2^2)A_1^2A_2^2\right]=0\end{aligned} \quad \ldots\ldots(27)$$

Hence

$$k_1A_1^2(\omega_1^2-\omega_{10}^2)=k_2A_2^2(\omega_2^2-\omega_{10}^2) \quad \ldots\ldots(28)$$

Substituting (28) and eliminating $\sin\phi$ and $\cos\phi$ from (21) through (24) one obtains

$$\frac{dA_1}{dt}=A_1\frac{\dfrac{\omega_2^2-\omega_{10}^2}{\omega_{10}^2+a_1^2}\left[\mu_1+\dfrac{\omega_1^2-\omega_{10}^2}{\omega_2^2-\omega_{10}^2}\cdot\dfrac{\omega_{10}^2+a_1^2}{\omega_{10}^2+a_2^2}\mu_2\right]}{2(1-k_1k_2)(\omega_{20}^2-\omega_{10}^2)} \quad \ldots\ldots(29)$$

$$\frac{d\phi_1}{dt}=\frac{\left[\mu_1a_1+\dfrac{\omega_1^2-\omega_{10}^2}{\omega_2^2-\omega_{10}^2}\cdot\dfrac{\omega_{10}^2+a_1^2}{\omega_{10}^2+a_2^2}\mu_2a_2\right]\dfrac{\omega_2^2-\omega_{10}^2}{\omega_{10}^2+a_1^2}}{2(1-k_1k_2)(\omega_{20}^2-\omega_{10}^2)} \quad \ldots\ldots(30)$$

Similarly considering the components of frequency ω_{20} it can be shown that

$$k_1B_1^2(\omega_1^2-\omega_{10}^2)=k_2B_2^2(\omega_2^2-\omega_{20}^2) \quad \ldots\ldots(31)$$

$$\frac{dB_1}{dt}=B_1\frac{\dfrac{\omega_2^2-\omega_{20}^2}{\omega_{20}^2+a_1^2}\left[\nu_1+\dfrac{\omega_1^2-\omega_{20}^2}{\omega_2^2-\omega_{20}^2}\cdot\dfrac{\omega_{20}^2+a_1^2}{\omega_{20}^2+a_2^2}\nu_2\right]}{2(1-k_1k_2)(\omega_{10}^2-\omega_{20}^2)} \quad \ldots\ldots(32)$$

$$\frac{d\theta_1}{dt}=\frac{\left[\nu_1a_1+\dfrac{\omega_1^2-\omega_{20}^2}{\omega_2^2-\omega_{20}^2}\cdot\dfrac{\omega_{20}^2+a_1^2}{\omega_{20}^2+a_2^2}\nu_2a_2\right]\dfrac{\omega_2^2-\omega_{20}^2}{\omega_{20}^2+a_1^2}}{2(1-k_1k_2)(\omega_{10}^2-\omega_{20}^2)} \quad \ldots\ldots(33)$$

Substituting the values of μ_1, μ_2, ν_1, ν_2 and putting

$$\frac{\omega_1^2-\omega_{10}^2}{\omega_2^2-\omega_{10}^2}=\alpha_1,\quad \frac{\omega_1^2-\omega_{20}^2}{\omega_2^2-\omega_{20}^2}=\alpha_2,\quad \frac{\omega_{10}^2+a_1^2}{\omega_{10}^2+a_2^2}=\beta_1 \quad\text{and}\quad \frac{\omega_{20}^2+a_1^2}{\omega_{20}^2+a_2^2}=\beta_2$$

$$\frac{\mathrm{d}A_1}{\mathrm{d}t}=A_1\frac{\omega_2^2-\omega_{10}^2}{\omega_{10}^2+a_1^2}\,\frac{[b_1-\frac{3}{4}c_1(A_1^2+2B_1^2)+\alpha_1\beta_1 b_2-\frac{3}{4}c_2\alpha_1\beta_1(A_2^2+2B_2^2)]}{2(1-k_1k_2)(\omega_{20}^2-\omega_{10}^2)} \qquad \text{......(34)}$$

$$\frac{\mathrm{d}B_1}{\mathrm{d}t}=B_1\frac{\omega_2^2-\omega_{20}^2}{\omega_{20}^2+a_1^2}\,\frac{[b_1-\frac{3}{4}c_1(B_1^2+2A_1^2)+\alpha_2\beta_2 b_2-\frac{3}{4}c_2\alpha_2\beta_2(B_2^2+2A_2^2)]}{2(1-k_1k_2)(\omega_{10}^2-\omega_{20}^2)} \qquad \text{......(35)}$$

Using (28) and (31) and putting

$$\frac{4}{3}\frac{b_1+\alpha_1\beta_1 b_2}{c_1+\alpha_1^2\frac{k_1}{k_2}\beta_1 c_2}=A_{10}^2,\qquad \frac{4}{3}\frac{b_1+\alpha_2\beta_2 b_2}{c_1+\alpha_2^2\frac{k_1}{k_2}\beta_2 c_2}=B_{10}^2$$

$$\frac{c_1+\alpha_1\beta_1\alpha_2\frac{k_1}{k_2}c_2}{c_1+\alpha_1^2\frac{k_1}{k_2}\beta_1 c_2}=\gamma_1,\qquad \frac{c_1+\alpha_2\beta_2\frac{k_1}{k_2}\alpha_1 c_2}{c_1+\alpha_2^2\frac{k_1}{k_2}\beta_2 c_2}=\gamma_2$$

$$\frac{\omega_2^2-\omega_{10}^2}{\omega_{10}^2+a_1^2}\cdot\frac{\frac{3}{4}\left(c_1+\alpha_1^2\frac{k_1}{k_2}\beta_2 c_2\right)}{2(1-k_1k_2)(\omega_{20}^2-\omega_{10}^2)}=\delta_1$$

$$\frac{\omega_2^2-\omega_{20}^2}{\omega_{20}^2+a_1^2}\cdot\frac{\frac{3}{4}\left(c_1+\alpha_2^2\frac{k_1}{k_2}\beta_2 c_2\right)}{2(1-k_1k_2)(\omega_{10}^2-\omega_{20}^2)}=\delta_2$$

one obtains

$$\frac{\mathrm{d}A_1}{\mathrm{d}t}=\delta_1A_1(A_{10}^2-A_1^2-2\gamma_1B_1^2) \qquad \text{......(36)}$$

$$\frac{\mathrm{d}B_1}{\mathrm{d}t}=\delta_2B_2(B_{10}^2-B_1^2-2\gamma_2A_1^2) \qquad \text{......(37)}$$

Equilibrium conditions are obtained on equating (36) and (37) to zero. The oscillations at the two frequencies may have the following combination of amplitudes

1. $A_1=A_{10}$, $\quad B_1=0$ ......(38)
2. $B_1=B_{10}$, $\quad A_1=0$ ......(39)
3. $A_1^2=\dfrac{2\gamma_1B_{10}^2-A_{10}^2}{4\gamma_1\gamma_2-1}$, $\quad B_1^2=\dfrac{2\gamma_2A_{10}^2-B_{10}^2}{4\gamma_1\gamma_2-1}$ ...(40)

The frequencies of oscillation are approximately equal to ω_{10} and ω_{20}. The actual frequency will differ from these values and the difference will be of the order of the coefficient of the non-linear term.

On applying Liapunoff's criteria[7] of stability it can be shown (Appendix 2) that

Combination 1 is stable if $\dfrac{A_{10}^2}{B_{10}^2}>\dfrac{1}{2\gamma_2}$ ......(41)

Combination 2 is stable if $\dfrac{B_{10}^2}{A_{10}^2}>\dfrac{1}{2\gamma_1}$ ......(42)

Combination 3 is stable if $2\gamma_1<\dfrac{A_{10}^2}{B_{10}^2}<\dfrac{1}{2\gamma_2}$ ......(43)

Thus, the oscillator system may oscillate in one mode at both the frequencies simultaneously and in another mode at either of the two frequencies separately. The two modes are mutually exclusive. Hence if it is arranged that $\dfrac{1}{2\gamma_2}<\dfrac{A_{10}^2}{B_{10}^2}<2\gamma_1$, the oscillator system may be used as a two-state device, each state being defined by a particular frequency of oscillation. The above characteristics are identical to those of the two-state oscillator consisting of two capacitively-coupled tuned-anode oscillators. However, in the present case the frequencies of oscillation are modified to the first order of the non-linear coefficient and the expressions for γ_1, γ_2 are more complicated.

2.3. *Circuit Parameters for Two-State Operation*

It has been shown by the analysis in the previous section that the present oscillator system may be used as a two-state device. Switching from one state to the other may be done by biasing off one of the valves. If it is considered that $\omega_1^2<\omega_2^2$ and the right-hand valve (V_r) is biased off, eqn (10) reduces to

$$\frac{\mathrm{d}^3V_2}{\mathrm{d}t^3}+a_2\frac{\mathrm{d}^2V_2}{\mathrm{d}t^2}+\omega_2^2\frac{\mathrm{d}V_2}{\mathrm{d}t}+a_2\omega_2^2V_2+ + k_1\frac{\mathrm{d}^3V_1}{\mathrm{d}t^3}+a_2k_1\frac{\mathrm{d}^2V_1}{\mathrm{d}t^2}=-\mu_2'V_1 \qquad \text{......(44)}$$

where
$$\mu_2'=\frac{1}{L_2C_2^2R_2}$$

Equations (36) and (37) are modified to

$$\frac{\mathrm{d}A_1}{\mathrm{d}t}=\delta_{1r}A_1(A_{1r}^2-A_1^2-2B_1^2) \qquad \text{......(45)}$$

$$\frac{\mathrm{d}B_1}{\mathrm{d}t}=\delta_{2r}B_1(B_{1r}^2-B_1^2-2A_1^2) \qquad \text{......(46)}$$

where
$$\delta_{1r}=\frac{c_1}{c_1+\alpha_1^2\frac{k_1}{k_2}\beta_2c_2}\delta_1 \qquad \text{......(47)}$$

$$\delta_{2r} = \frac{c_1}{c_1 + \alpha_2^2 \dfrac{k_1}{k_2} \beta_2 c_2} \quad \ldots\ldots(48)$$

$$A_{1r}^2 = \frac{4}{3c_1}(b_1 - \alpha_1 \beta_1 \mu_2') \quad \ldots\ldots(49)$$

$$B_{1r}^2 = \frac{4}{3c_1}(b_1 - \alpha_2 \beta_2 \mu_2') \quad \ldots\ldots(50)$$

Since ω_1^2 is assumed to be less than ω_2^2, α_1 is less than α_2 and it can be arranged that

$$2(b_1 - \alpha_2 \beta_2 \mu_2') > (b_1 - \alpha_1 \beta_1 \mu_2') \quad \ldots\ldots(51)$$

The combination (1) is only stable with the above adjustment and the oscillator system is forced to oscillate at ω_{10} independent of its previous state. On removing the bias, after the oscillator has switched to the ω_{10} state, it will stay in this state.

Similarly switching from the ω_{10} state to ω_{20} state may be effected by biasing off the left-hand side valve, and putting it on after the switching time. The inequality to be satisfied is

$$(\alpha_1 \beta_1 b_2 - \mu_1') < 2(\alpha_2 \beta_2 b_2 - \mu_1') \quad \ldots\ldots(52)$$

where $$\mu_1' = \frac{1}{L_1 C_1^2 R_1}$$

It is then evident that to operate the oscillator system as a two-state device, the parameters are to be so chosen that inequalities (51), (52), (41) and (42) are satisfied.

3. An Experimental Oscillator System

The circuit arrangement of an oscillator system constructed for experimental verification of the analysis presented in the earlier section is shown in Fig. 2. The circuit parameters were adjusted to satisfy the inequalities (41), (42), (51) and (52) by adjusting the values of M, the tuning capacitances and the load resistances. Operation of the oscillator system at either of the two frequencies as well as switching from one state to the other was possible. The switching time for frequencies of oscillation near about 20 Mc/s was of the order of 1–2 μs. The power consumption was approximately 1 watt. Photographs showing the switching from one state to the other are given in Fig. 3. Simultaneous oscillation at the two frequencies were exhibited for certain parameters of the oscillator system agreeing approximately with the inequality (43).

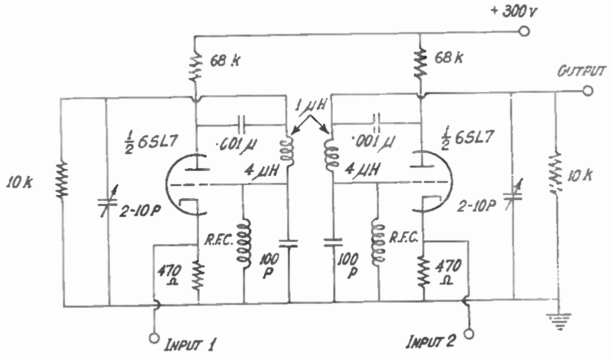


Fig. 2. Circuit diagram of the experimental oscillator system.

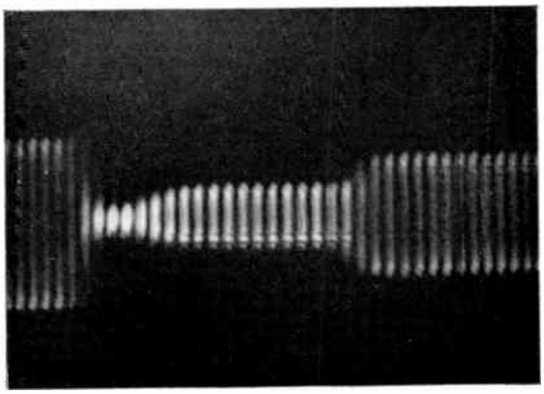


Fig. 3. Photograph showing the switching from one state to the other. (Intensity markers are at intervals of 1 μs.)

It may be mentioned that the switching time realizable by conventional circuits using the same valves is of the order of 10 μs.[8] The system may, therefore, be used as a two-state device allowing a better speed of operation, and also providing non-destructive output in both the states.

4. Conclusions

The analysis of an oscillator system consisting of two inductively-coupled Colpitts oscillators show that the system may be used as a two-state device. Switching from one state to the other may be achieved by biasing off one of the tubes. In this Colpitts form, since the frequencies of oscillation may be increased to the limit permitted by the tube parameters, significant improvement in the resolving time over conventional circuits may be achieved. These results have been confirmed approximately by experiments.

5. Acknowledgment

The author is deeply indebted to Professor J. N. Bhar, D.Sc., F.N.I., for his kind interest in the work.

6. References

1. W. A. Edson, "Frequency memory in multi-mode oscillators", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-2**, p. 58, March 1955.
2. M. I. Disman, "Registers and Counters based on Frequency Memory", Technical Report No. 19, 16th August 1954, Electronics Research Laboratory, Stanford University.

3 B. van der Pol, "On oscillation hysteresis in a triode generator with two degrees of freedom", *Philosophical Magazine*, **43**, p. 700, 1922.

4. R. E. Fontana, "Internal resonance in circuits containing non-linear resistance", *Proc. Inst. Radio Engrs*, **39**, p. 945, 1951.

5. L. V. Skinner, "Criteria of Stability in Circuits Containing Nonlinear Resistance", Doctoral Dissertation, University of Illinois, 1948.

6. J. S. Schaffner, "Simultaneous oscillation in oscillators", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-1**, p. 2, 1954.

7. I. Liapunov, "Problème Général de la Stabilité du mouvement", Reprinted in 1949 by Princeton University Press.

8. N. C. Elmore and M. Sands, "Electronics", p. 209. (McGraw Hill, New York, 1949.)

7. Appendix 1

Equations (21)–(24) are obtained in the following way:

Neglecting the harmonics and combination frequencies, V_1^3 and V_2^3 may be written as

$$V_1^3 = \tfrac{3}{4}A_1(A_1^2+2B_1^2)\sin(\omega_{10}t+\phi_1)+\tfrac{3}{4}B_1(B_1^2+2A_1^2)\sin(\omega_{20}t+\theta_1) \quad \ldots\ldots(53)$$

$$V_2^3 = \tfrac{3}{4}A_2(A_2^2+2B_2^2)\sin(\omega_{10}t+\phi_2)+\tfrac{3}{4}B_2(B_2^2+2A_2^2)\sin(\omega_{20}t+\theta_2) \quad \ldots\ldots(54)$$

Right-hand side of (15) and (16) may be reduced to

$$-b_1 V_1+c_1 V_1^3 = \mu_1 A_1\sin(\omega_{10}t+\phi_1)+\nu_1 B_1\sin(\omega_{20}t+\theta_1) \quad \ldots\ldots(55)$$

$$-b_2 V_2+c_2 V_2^3 = \mu_2 A_2\sin(\omega_{10}t+\phi_2)+\nu_2 B_2\sin(\omega_{20}t+\theta_2) \quad \ldots\ldots(56)$$

where

$$\mu_1 = \tfrac{3}{4}c_1(A_1^2+2B_1^2)-b_1 \quad \ldots\ldots(57) \qquad \mu_2 = \tfrac{3}{4}c_2(A_2^2+2B_2^2)-b_2 \quad \ldots\ldots(59)$$

$$\nu_1 = \tfrac{3}{4}c_1(B_1^2+2A_1^2)-b_1 \quad \ldots\ldots(58) \qquad \nu_2 = \tfrac{3}{4}c_2(B_2^2+2A_2^2)-b_2 \quad \ldots\ldots(60)$$

Henceforward the two frequency components of (17) and (18) can be considered separately. Differentiating (17)

$$\frac{dV_1}{dt} = \omega_{10}A_1\cos(\omega_{10}t+\phi_1)+\frac{dA_1}{dt}\sin(\omega_{10}t+\phi_1)+A_1\frac{d\phi_1}{dt}\cos(\omega_{10}t+\phi_1) \quad \ldots\ldots(61)$$

Differentiating again and neglecting $\frac{d^2A_1}{dt^2}, \frac{d^2\phi_1}{dt^2}, \frac{dA_1}{dt}\cdot\frac{d\phi_1}{dt}$

$$\frac{d^2V_1}{dt^2} = -\omega_{10}^2A_1\sin(\omega_{10}t+\phi_1)+2\omega_{10}\frac{dA_1}{dt}\cos(\omega_{10}t+\phi_1)-2\omega_{10}A_1\frac{d\phi_1}{dt}\sin(\omega_{10}t+\phi_1) \quad \ldots\ldots(62)$$

Similarly, differentiating successively and neglecting the higher order derivatives and products of the derivatives

$$\frac{d^3V_1}{dt^3} = -\omega_{10}^3A_1\cos(\omega_{10}t+\phi_1)-3\omega_{10}^2\frac{dA_1}{dt}\sin(\omega_{10}t+\phi_1)-3\omega_{10}^2A_1\frac{d\phi_1}{dt}\cos(\omega_{10}t+\phi_1) \quad \ldots\ldots(63)$$

$$\frac{d^4V_1}{dt^4} = \omega_{10}^4A_1\sin(\omega_{10}t+\phi_1)-4\omega_{10}^3\frac{dA_1}{dt}\cos(\omega_{10}t+\phi_1)+4\omega_{10}^3A_1\frac{d\phi_1}{dt}\sin(\omega_{10}t+\phi_1) \quad \ldots\ldots(64)$$

$$\frac{d^5V_1}{dt^5} = \omega_{10}^5A_1\cos(\omega_{10}t+\phi_1)+5\omega_{10}^4\frac{dA_1}{dt}\sin(\omega_{10}t+\phi_1)+5\omega_{10}^4A_1\frac{d\phi_1}{dt}\cos(\omega_{10}t+\phi_1) \quad \ldots\ldots(65)$$

$$\frac{d^6V_1}{dt^6} = -\omega_{10}^6A_1\sin(\omega_{10}t+\phi_1)+6\omega_{10}^5\frac{dA_1}{dt}\cos(\omega_{10}t+\phi_1)-6\omega_{10}^5A_1\frac{d\phi_1}{dt}\sin(\omega_{10}t+\phi_1) \quad \ldots\ldots(66)$$

Similarly, one obtains

$$\frac{d^2V_2}{dt^2} = -\omega_{10}^2A_2\sin(\omega_{10}t+\phi_2)+2\omega_{10}\frac{dA_2}{dt}\cos(\omega_{10}t+\phi_2)-2\omega_{10}A_2\frac{d\phi_2}{dt}\sin(\omega_{10}t+\phi_2) \quad \ldots\ldots(67)$$

$$\frac{d^3V_2}{dt^3} = -\omega_{10}^3A_2\cos(\omega_{10}t+\phi_2)-3\omega_{10}^2\frac{dA_2}{dt}\sin(\omega_{10}t+\phi_2)-3\omega_{10}^2A_2\frac{d\phi_2}{dt}\cos(\omega_{10}t+\phi_2) \quad \ldots\ldots(68)$$

Combining (15) and (16)

$$(1-k_1 k_2)\frac{d^6V_1}{dt^6}+(1-k_1 k_2)(a_1+a_2)\frac{d^5V_1}{dt^5}+\{\omega_1^2+\omega_2^2+(1-k_1 k_2)a_1 a_2\}\frac{d^4V_1}{dt^4}+(a_1+a_2)(\omega_1^2+\omega_2^2)\frac{d^3V_1}{dt^3}+$$
$$+\{\omega_1^2\omega_2^2+a_1 a_2(\omega_1^2+\omega_2^2)\}\frac{d^2V_1}{dt^2}+(a_1+a_2)\omega_1^2\omega_2^2\frac{dV_1}{dt}+a_1 a_2\omega_1^2\omega_2^2 V_1$$
$$=\left(\frac{d^3}{dt^3}+a_2\frac{d^2}{dt^2}+\omega_2^2\frac{d}{dt}+a_2\omega_2^2\right)(-b_1 V_1^2+C_1 V_1^3)-k_2\left(\frac{d^3}{dt^3}+a_1\frac{d^2}{dt^2}\right)(-b_2 V_2+C_2 V_2^3)\text{......(69)}$$

Substituting in (17), (55), (56) and (61) to (68) and noting that

$$\omega_{10}^4(1-k_1 k_2)-\omega_{10}^2(\omega_1^2+\omega_2^2)+\omega_1^2\omega_2^2=0$$

one obtains

$$(a_1+a_2)[5\omega_{10}^4(1-k_1k_2)-3(\omega_1^2+\omega_2^2)\omega_{10}^2+\omega_1^2\omega_2^2]\frac{dA_1}{dt}\sin(\omega_{10}t+\phi_1)+$$
$$+[6\omega_{10}^5(1-k_1k_2)-4\omega_{10}^3\{\omega_1^2+\omega_2^2+a_1a_2(1-k_1k_2)\}+2\omega_{10}\{\omega_1^2\omega_2^2+a_1a_2(\omega_1^2+\omega_2^2)\}]\frac{dA_1}{dt}\cos(\omega_{10}t+\phi_1)+$$
$$+(a_1+a_2)[5\omega_{10}^4(1-k_1k_2)-3(\omega_1^2+\omega_2^2)\omega_{10}^2+\omega_1^2\omega_2^2]A_1\frac{d\phi_1}{dt}\cos(\omega_{10}t+\phi_1)-$$
$$-[6\omega_{10}^5(1-k_1k_2)-4\omega_{10}^3\{\omega_1^2+\omega_2^2+a_1a_2(1-k_1k_2)\}+2\omega_{10}\{\omega_1^2\omega_2^2+a_1a_2(\omega_1^2+\omega_2^2)\}]A_1\frac{d\phi_1}{dt}\sin(\omega_{10}t+\phi_1)$$
$$=(\omega_2^2-\omega_{10}^2)\omega_{10}\mu_1A_1\cos(\omega_{10}t+\phi_1)+a_2(\omega_2^2-\omega_{10}^2)\mu_1A_1\sin(\omega_{10}t+\phi_1)-$$
$$-k_2\omega_{10}^3\mu_2A_2\cos(\omega_{10}t+\phi_2)-k_1a_1\omega_{10}^2\mu_2A_2\sin(\omega_{10}t+\phi_2)\text{......(70)}$$

Equation (70) is written neglecting the products of μ_1, μ_2 and the derivative of the amplitudes and phases, since the products have magnitudes of second order.

Equating the sine and cosine terms on the two sides of (70), noting that

$$5\omega_{10}^4(1-k_1 k_2)-3(\omega_1^2+\omega_2^2)\omega_{10}^2+\omega_1^2\omega_2^2=2\omega_{10}^2(1-k_1 k_2)(\omega_{10}^2-\omega_{20}^2)\quad\text{......(71)}$$

$$6\omega_{10}^5(1-k_1 k_2)-4\omega_{10}^3\{\omega_1^2+\omega_2^2+a_1 a_2(1-k_1 k_2)+2\omega_{10}\{\omega_1^2\omega_2^2+a_1 a_2(\omega_1^2+\omega_2^2)\}$$
$$=2\omega_{10}(1-k_1 k_2)(\omega_{10}^2-\omega_{20}^2)(\omega_{10}^2-a_1 a_2)\text{......(72)}$$

and substituting

$$2\omega_{10}(1-k_1 k_2)(\omega_{10}^2-\omega_{20}^2)=n\quad\text{......(73)}$$

one obtains

$$\omega_{10}n(a_1+a_2)\frac{dA_1}{dt}-n(\omega_{10}^2-a_1 a_2)A_1\frac{d\phi_1}{dt}$$
$$=a_2(\omega_2^2-\omega_{10}^2)\mu_1 A_1+k_2\omega_{10}^2\mu_2 A_2(\omega_{10}\sin\phi-a_1\cos\phi)\quad\text{......(74)}$$

$$n(\omega_{10}^2-a_1 a_2)\frac{dA_1}{dt}+\omega_{10}n(a_1+a_2)A_1\frac{d\phi_1}{dt}$$
$$=\omega_{10}(\omega_2^2-\omega_{10}^2)\mu_1 A_1-k_2\omega_{10}^2\mu_2 A_2(\omega_{10}\cos\phi+a_1\sin\phi)\quad\text{......(75)}$$

where $\phi=\phi_2-\phi_1$

Similarly, combining (15) and (16) and writing the left-hand side in terms of V_2 one obtains

$$\omega_{10}n(a_1+a_2)\frac{dA_2}{dt}-n(\omega_{10}^2-a_1 a_2)A_2\frac{d\phi_2}{dt}$$
$$=a_1(\omega_1^2-\omega_{10}^2)\mu_2 A_2-k_1\omega_{10}^2\mu_1 A_1(\omega_{10}\sin\phi+a_2\cos\phi)\quad\text{......(76)}$$

$$n(\omega_{10}^2-a_1 a_2)\frac{dA_2}{dt}+n\omega_{10}(a_1+a_2)A_2\frac{d\phi_2}{dt}$$
$$=\omega_{10}(\omega_1^2-\omega_{10}^2)\mu_2 A_2-k_1\omega_{10}^2\mu_1 A_1(\omega_{10}\cos\phi-a_2\sin\phi)\quad\text{......(77)}$$

8. Appendix 2

The inequalities that the equilibrium values must satisfy for stability are[7]

$$\left[\frac{\partial}{\partial A_1}\left(\frac{\partial A_1}{\partial t}\right)\right]_0 + \left[\frac{\partial}{\partial B_1}\left(\frac{\partial B_1}{\partial t}\right)\right]_0 < 0 \quad \ldots\ldots(78)$$

$$\left[\frac{\partial}{\partial A_1}\left(\frac{\partial A_1}{\partial t}\right)\right]_0 \left[\frac{\partial}{\partial B_1}\left(\frac{\partial B_1}{\partial t}\right)\right]_0 - \\ - \left[\frac{\partial}{\partial A_1}\left(\frac{\partial B_1}{\partial t}\right)\right]_0 \left[\frac{\partial}{\partial B_1}\left(\frac{\partial A_1}{\partial t}\right)\right]_0 > 0 \quad \ldots\ldots(79)$$

On substituting for the derivatives from (35) and (36) one obtains

$$\delta_1(A_{10}^2 - 3A_1^2 - 2\gamma_1 B_1^2) + \delta_2(B_{10}^2 - 3B_1^2 - 2\gamma_2 A_1^2) < 0 \quad \ldots\ldots(80)$$

$$\delta_1 \delta_2 (A_{10}^2 - 3A_1^2 - 2\gamma_1 B_1^2)(B_{10}^2 - 3B_1^2 - 2\gamma_2 A_1^2) - \\ - \delta_1 \delta_2 16\gamma_1 \gamma_2 A_1^2 B_1^2 > 0 \quad \ldots\ldots(81)$$

Substituting the equilibrium values from (38) one obtains

$$-\delta_1(2A_{10}^2) + \delta_2(B_{10}^2 - 2\gamma_2 A_{10}^2) < 0 \quad \ldots\ldots(82)$$

$$-2A_{10}^2(B_{10}^2 - 2\gamma_2 A_{10}^2) > 0 \quad \ldots\ldots(83)$$

The two inequalities are satisfied if

$$\frac{A_{10}^2}{B_{10}^2} > \frac{1}{2\gamma_2} \quad \ldots\ldots(84)$$

Proceeding, similarly combination (2) is stable if

$$\frac{B_{10}^2}{A_{10}^2} > \frac{1}{2\gamma_1} \quad \ldots\ldots(85)$$

Substituting the equilibrium values from (40) in (80) and (81) one obtains

$$-2\delta_1 A_1^2 - 2\delta_2 B_1^2 < 0 \quad \ldots\ldots(86)$$

$$1 - 4\gamma_1 \gamma_2 > 0 \quad \ldots\ldots(87)$$

Inequality (87) is satisfied if $\gamma_1\gamma_2 < \frac{1}{4}$. In addition it is required that A_1^2 and B_1^2 given by (40) should be positive. This requires that A_{10}^2 and B_{20}^2 should satisfy the following inequalities

$$2\gamma_1 < \frac{A_{10}^2}{B_{10}^2} < \frac{1}{2\gamma_2} \quad \ldots\ldots(88)$$

Since inequality (87) is satisfied if (88) is satisfied, combination (3) is stable if inequality (88) is satisfied.

Manuscript received by the Institution on 28th August 1961 (Paper No. 739).

U.D.C. 621.391.812.62.029.62/4

Propagation Influences in Microwave Link Operation

By

M. W. GOUGH, M.A.†

Presented at a meeting of the South Midlands Section in Malvern on 29th March 1960.

Summary: Basic characteristics of ground-to-ground propagation in the frequency band 50–10 000 Mc/s are discussed, with emphasis on the operation of point-to-point radio links. Practical applications of such knowledge are illustrated by appraisals of the use of flat radio mirrors for circumventing mountain obstructions on microwave radio links, and of the capabilities of large-aperture aerials and space-diversity systems in reducing fading. Ground reflections are shown to be a key factor in the performance of very short wave radio links and the influence of inhomogeneities in the lower atmosphere is also stressed. It is pointed out how statistical analysis of temporal and spatial signal strength variations can often reveal the nature of the propagation mechanism at work, while at the same time pointing the way to improvements in circuit reliability.

List of Symbols

ρ	reflection coefficient for rough surface.
ρ_s	reflection coefficient for smooth surface.
λ	wavelength.
δ	mean height of surface irregularities.
θ	ray reflection angle.
θ_c	critical ray angle for total internal reflection.
D	path length.
H	virtual height of radio terminal.
S	half the vertical interference-fringe distance.
f	vertical interference-fringe distance.
a	radio-mirror or aerial aperture.
a/f	aperture-fringe ratio.
d	horizontal interference-fringe distance.
α	horizontal angle between direct and laterally reflected ray paths.
d_1	mirror distance from nearer radio terminal.
d_2	mirror distance from further radio terminal.
A_p	projected mirror area.
a_p	projected mirror dimension.
l	mirror length.
d_{min}	minimum permissible separation between mirror centres.
β	angle of deviation between incident and mirror-reflected ray paths.
n	atmospheric refractive index at a height h above the earth.
dn/dh	refractive index gradient.
R_0	true earth radius (3960 miles).
M	modified refractive index at a height h above the earth $= 10^6[n-1+h/R_0]$
dM/dh	modified refractive index gradient $= 10^6[dn/dh+1/R_0]$
K	effective earth radius factor $= \dfrac{1}{R_0 dn/dh+1} = \dfrac{10^6}{R_0 dM/dh}$
ΔM	modified refractive index deficit.
W_1	track width for first propagation mode.
W_2	track width for second propagation mode.
W_3	track width for third propagation mode.
J_1	Bessel function of first order.

1. Introduction

This paper sets out to review some of the more impressive—and sometimes at first sight surprising and curious—features of radio wave propagation in the frequency band 50–10 000 Mc/s, with particular reference to point-to-point links. No pretence is made to a comprehensive or balanced survey of the extensive literature concerned with this field of enquiry. Instead, a critical analysis has been attempted of selected microwave measurements (described in the literature or connected with the author's own work), designed firstly to display peculiarities in propagation behaviour to which conventional radio links are subject, and secondly to illuminate some of the funda-

† Marconi's Wireless Telegraph Co. Ltd., Baddow Research Laboratories, Great Baddow, Essex.

mental processes of microwave propagation, the understanding of which assists exploitation of the transmission medium to best advantage while pointing the way to possible improvements in communications technique. This appraisal is linked with assessments of the potentialities of passive repeaters and diversity techniques, both of which are likely to find increasing application in difficult situations arising on microwave routes. While some of the conditions described are admittedly extreme, they are all associated with commercial installations, many of which are now in operation.

The term "microwave" has been used to embrace the whole frequency band under discussion. Indeed, this wide spectrum has some claim to a single designation, in that throughout its range under "point-to-point" conditions there is virtually no ionospheric influence, the effect of surface dielectric properties is subordinate, atmospheric absorption is small and—of paramount importance—propagation behaviour is very susceptible to small changes in the atmospheric refractive index.

2. Line-of-Sight Paths and the Characteristics of Surface Reflections

Above about 400 Mc/s, a line-of-sight path is usually essential for obtaining an acceptable performance from a conventional point-to-point radio link, but whatever the frequency employed, we must come to terms with the influence of the *ground*. Disregarding for the present the effect of the atmosphere and of intervening obstructions, we are faced in the simplest situation with interference between direct and ground-reflected radio paths. The relative amplitude and phase of significant ground reflections often provide the key to the behaviour of very short wave radio paths. In the v.h.f. band these factors affect the mean signal strength, while at higher frequencies they usually control the severity of fading. An appreciation of the reflecting properties of land and water surfaces is thus of paramount importance in engineering very short wave radio systems. A study of the subject reveals a surprising number of subtleties, however.

2.1. *Fresnel Reflection Coefficients*

Above about 50 Mc/s, the Fresnel reflection coefficient for smooth ground is virtually independent of frequency and ground conductivity, depending mainly on the wave polarization, the ground dielectric constant and the reflection angle. For vertical polarization there is a Brewster effect (familiar in optics), whereby the coefficient becomes zero at a critical reflection angle, which for land varies between about 15 and 25 degrees according to its dielectric constant. Although the Brewster effect always ensures a smaller reflection coefficient with vertical than with horizontal polarization, for land the difference is usually insignificant at the shallow reflection angles characteristic of optical ground-to-ground propagation. Thus the ground irregularities, rather than the polarization, exert a dominant influence in the frequency band under review.

For sea, the reflection coefficient for vertical polarization has a minimum (but not zero) value at the pseudo-Brewster angle which, although itself dependent on the wave-length, is always much shallower than its Brewster counterpart for land. The resulting reduction in reflection coefficient for shallow angles as compared with horizontal polarization usefully limits the severity of destructive interference from sea reflections. The benefit from vertical polarization declines, however, with increasing frequency.

Reflection coefficients for land and sea, which involve many complexities omitted here, have been displayed graphically for radio waves by Kerr[1] and others.

Vertical polarization gives a smaller reflection coefficient than horizontal, and this fact favours the former whenever the signal strength is likely to be weakened by phase opposition between direct and surface-reflected waves. This condition often occurs in the u.h.f. and s.h.f. bands when the terminals are sited at a considerable height above the reflecting sea, thereby involving a large path difference in terms of wavelength between direct and reflected waves. The advantage of vertical polarization in these circumstances has been amply confirmed by Quarta[2] during tests on 1000 Mc/s over an optical sea path in North Italy. Figure 1 shows the cyclic field strength patterns obtained when the receiving aerial at Portofino was raised through a height interval of 65 ft while the transmitting terminal remained fixed. These

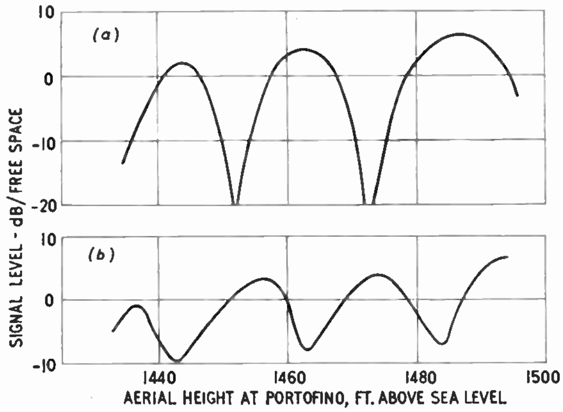


Fig. 1. Effect of polarization on 1000 Mc/s oversea path in Italy. (After Quarta.[2])

(*a*) Horizontal polarization. (*b*) Vertical polarization.

interference patterns, of a type referred to subsequently, result from the progressive phase shift between the two ray paths engendered by raising one of the aerials. Figure 1 (*a*), obtained with horizontal polarization, shows deep nulls indicative of specular sea reflection while Fig. 1 (*b*), secured with vertical polarization, shows much shallower minima, testifying to the reduced reflection coefficient characteristic of that polarization. The upward trend of the experimental curves is due to diffraction at some 40 ft trees near the moving aerial.

2.2. *Influence of Ground Roughness*

In planning microwave links it is important to select routes where the ground reflection coefficient does not exceed 0·5, in order to avoid undesirable interference effects manifest as fading during atmospheric changes. Fortunately under many practical conditions the reflected wave is scattered by ground roughness to a useful extent. The subject has been comprehensively treated by Bachynski.[3]

Rayleigh's well-known principle, originally enunciated for light waves, relates the diffuseness of reflection to the quantity $\delta \sin \theta/\lambda$, where, as indicated on Fig. 2, δ is the height of the irregularities above and

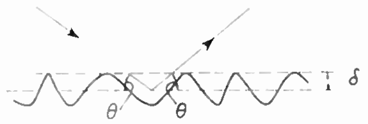


Fig. 2. Reflection at a rough surface.

below the mean level of a rough surface, θ is the reflection angle, and λ is the wavelength. It is to be expected that for rough surfaces of similar statistical character, there will be a unique relationship between $\delta \sin \theta/\lambda$ and the ratio ρ/ρ_s, where ρ and ρ_s are respectively the reflection coefficients of the surface when rough and when smooth.

In an attempt to discover an empirical relationship for radio waves, the author has collated measurements of over-land reflection coefficients from various sources,[4, 5] choosing situations where the representative ground roughness δ and the reflection angle θ were approximately known. Figure 3 shows the resulting plot of the ratio ρ/ρ_s against the quantity $\delta \sin \theta/\lambda$.

Court[6] has discovered empirically that there is an approximately exponential relationship between the reflection coefficient and the reflection angle for monochromatic light reflected diffusely at a matt surface. Bearing in mind Rayleigh's principle, an analogous exponential relationship has been devised to give the best fit to the present plot of radio reflection

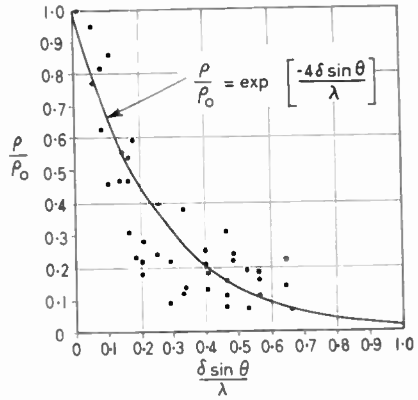


Fig. 3. Empirical relation between reflection coefficient and ground roughness.

coefficients. The resulting empirical relationship is given by

$$\rho/\rho_s = \exp(-4\delta \sin \theta/\lambda) \qquad \ldots\ldots(1)$$

This relationship is superimposed on Fig. 3. Putting $\rho_s = 1$, which is justifiable for over-land paths, the important condition that $\rho < 0{\cdot}5$ requires

$$\delta \sin \theta/\lambda > 0{\cdot}17 \text{ approx.} \qquad \ldots\ldots(2)$$

For a typical optical path this condition requires ground irregularities exceeding about ± 30 ft at 3000 Mc/s and ± 600 ft at 150 Mc/s. Thus, while in the s.h.f. band the reflection coefficient is often kept low by natural ground irregularities, in the v.h.f. band quite hilly terrain can reflect efficiently. Indeed, ground reflection is usually inseparable from ground-to-ground v.h.f. propagation.

In contrast to point-to-point links, radar operation often exploits reflecting ground in front of the equipment for generating the requisite aerial lobe pattern. Norton and Omberg[7] state that a well-developed lobe structure requires the reflecting ground to be smooth enough to satisfy

$$\delta < \frac{\lambda}{16 \sin \theta} \qquad \ldots\ldots(3)$$

where δ and θ are as defined in Fig. 2.

Reference to Fig. 3 shows this rule to imply the existence of an adequate lobe pattern when $\rho/\rho_s > 0{\cdot}8$.

2.3. "*Height–Gain*" *Measurements*

"Height–gain" measurements comprise a powerful means of assessing the reflecting properties of the ground, and are of particular assistance in predicting the behaviour of s.h.f. links, whose performance can be severely impaired both by surface reflections and by obstructions. These tests involve the measurement (during a period of stable propagation) of the variations in received signal level resulting from progressively raising one of the terminals while the

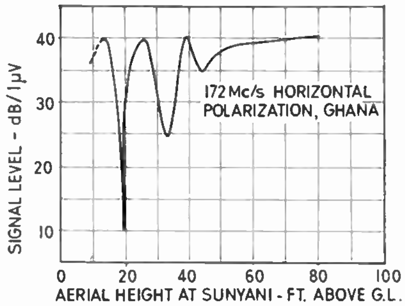


Fig. 4. Height–gain curve showing decline of ground-reflection interference.

other remains fixed at some suitable height. For an intervening flat reflecting surface this leads to a succession of nearly equally spaced interference fringes. Furthermore the reflection coefficient of the ground can be computed from measurement of adjacent peak and trough levels. As the moving aerial is raised the reflecting region recedes from it, with the result that over an irregular or curved path the character of the fringes can alter with height as the reflecting region changes. Figure 4 shows a noteworthy interference pattern obtained in Ghana on a frequency of 172 Mc/s. The great depth of the first null is attributable to a specular reflection at a steep slope in the foreground of the moving aerial. The reflection quickly dies out however as increased aerial height moves the first Fresnel zone† to the bottom of the hill, where it is progressively curtailed by the ensuing flat ground. The chart affords a good example of the catastrophic impairment in performance that may occur if a hill-top terminal, visually well sited, utilizes an aerial that by ill fortune is mounted in the neighbourhood of a deep signal strength null. It is to guard against such expensive accidents that exploratory radio surveys are usually carried out before the installation of permanent point-to-point links.

Figure 5, after Day and Trolese,[5] shows height–gain curves obtained on three microwave frequencies tested over the Arizona desert, using an aircraft ranging in height between 600 and 7700 ft at a distance of about 55 miles from the transmitter. Inspection of the depth of the nulls in the interference patterns shows the generally increasing diffuseness of the ground reflection with increasing frequency, in broad conformity with Rayleigh's principle already discussed. However at 3300 Mc/s the height–gain pattern has become incoherent, with sporadic occurrences of high reflection coefficients. These latter represent "glints" off randomly angled surfaces (possibly large rocks) that are sufficiently extensive to be comparable with the first Fresnel zone at this high frequency, and thereby become reflectors in their own right. In this situation Rayleigh's principle is not applicable because the surface can no longer be treated statistically.

† The first Fresnel zone for reflection at a flat surface, which for grazing angles is bounded by an elongated ellipse surrounding the point of geometrical reflection, may broadly be considered as defining the minimum surface capable of reflecting as effectively as a flat surface of infinite extent. Thus a smaller flat surface, restricted for example by a sudden change of slope or other discontinuity, will reflect progressively less as its area diminishes.

3. Some Properties of Obstructed Paths

Fortunately for the radio engineer, radio wave diffraction round the earth's surface and over hills and other obstructions often provides sufficient signal strength for operating point-to-point links in the v.h.f. band. The ability of these frequencies to give a useful signal within the "shadow" of obstructions, as exhibited by application of Fresnel's edge-diffraction theory, is a particularly valuable feature in inaccessible country where the exclusive use of optical paths would greatly increase the radio mileage. However, in the u.h.f. and s.h.f. bands obstruction losses can rarely be tolerated, and it is necessary to secure at least first Fresnel zone clearance of all obstacles along the path. Exploratory height–gain measurements over the path are very useful in determining the minimum aerial heights satisfying this condition, and also in locating salient obstacles, such as buildings, trees and hills, that are impairing performance at lower heights.

3.1. *Obstacle Gain*

A surprising and valuable by-product of the classical theories of "knife-edge" and "curved earth" diffraction is their implication of the phenomenon of

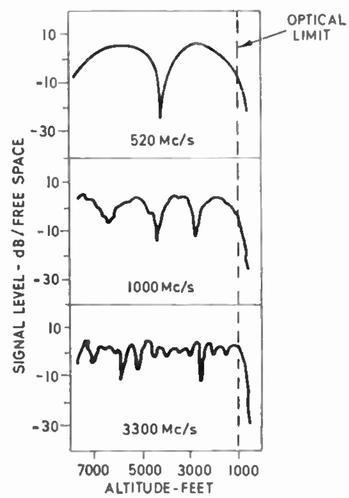
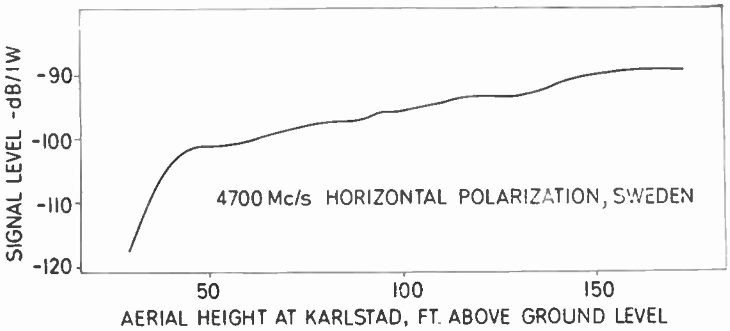


Fig. 5. Aircraft height–gain curves. (After Day and Trolese.[5])

Fig. 6.
Height–gain curve showing influence of near and distant obstructions.



"obstacle gain", whereby the presence of a suitable obstruction on a long non-optical path can engender a much greater field than is obtainable in the obstacle's absence. The "obstacle gain" is defined simply as the improvement in signal level (in decibels) brought about by the intervention of the obstacle. Necessary conditions for achieving a high obstacle gain are, firstly, that the terminals should be non-optical in the obstacle's absence (thereby suffering a high attenuation due to curved earth diffraction), and, secondly, that the top of the obstacle should be visible from each terminal. Neglecting possible ground reflections and inadequate ground clearance between the obstacle and the respective terminals, the path will give a signal strength falling short of the "free space" value by the calculable amount of the edge-diffraction loss. This signal level can far exceed that applicable purely to "curved earth" diffraction.

Striking confirmation of obstacle gain has been reported by Dickson and others,[8] after tests on 38 Mc/s over a 160 mile path in Alaska centrally obstructed by Mount Fairweather, an 8775 ft peak. With aerial heights of about 200 ft the path gave an attenuation close to the theoretical figure. The measured obstacle gain was 73 dB, with signal strengths varying by only ± 2 dB during 30 days of test.

3.2. *Microwave "Height–Gain" Tests as an Index of Obstructions*

The influence of obstructions along a radio path is often clearly revealed by the height–gain tests that form part of preliminary surveys designed to establish the reliability of prospective radio links. Figure 6 shows such a height–gain curve obtained in Sweden, using a frequency of 4700 Mc/s. The initial steep slope of the curve results from severe shadowing by neighbouring 40 ft trees. Once the line of sight has cleared these trees by a few feet their influence becomes negligible, and there follows a slower improvement in signal strength controlled by an obstructing hill 6 miles away, which itself is subsequently cleared at an aerial height of about 160 ft. At greater heights virtually "free space" conditions prevail.

4. Multi-path Propagation, with particular reference to Lateral Reflections

The existence of more than one significant radio path over a point-to-point link can increase inter-channel crosstalk in frequency-division-multiplex multichannel systems, to a progressive degree as the number of voice channels is increased. Impairments in the performance of installed high-capacity microwave systems attributable to multi-path propagation are rumoured from time to time, but concrete facts are difficult to discover in the literature. Fabbri,[9] after examining the performance of 22 such paths in Italy, mostly operating on 6000 Mc/s, concludes that, while 120 channels are unlikely to suffer significantly from multi-path effects, 300 channels require particularly favourable conditions to avoid appreciable impairment. Fabbri emphasizes however that this is not an ultimate limit. He predicts further improvements with narrower beams obtainable either by the use of larger aerials or higher frequencies.

A general discussion of multi-path propagation, devoid of experimental backing, would be out of place here. Instead a brief description will be given of an unexpected observation by the author of multi-path by lateral reflection, in a situation that greatly simplified its detection.

4.1. *Multi-path Propagation by Lateral Reflection at V.H.F.*

Figure 7 shows the region involved in propagation over a 12-mile radio path in Shropshire, proposed by the Midlands Electricity Board as a v.h.f. link in connection with a radio-operated power line protection system. The test transmitter was at Comet Bridge, a sub-station near Shrewsbury, while the receiving aerial could be sited anywhere on the boiler house roof of the Central Electricity Generating Board's station situated at Ironbridge in the Severn Valley. This route suffered the disability that the "direct" path TR was severely obstructed by a shoulder of the Wrekin, resulting in a weak diffracted signal. An investigation of discrepancies in an initial series of signal strength measurements eventually led

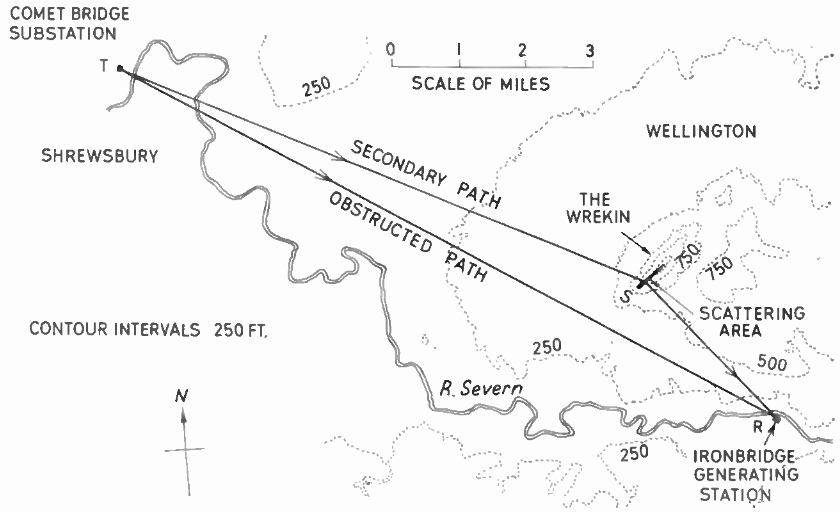


Fig. 7. Region involved in v.h.f. multi-path propagation.

to the discovery of pronounced horizontal interference fringes at the Ironbridge receiving terminal.

Figure 8 shows the interference patterns at the proposed respective operating frequencies of 173·4 and 183·4 Mc/s, obtained by moving the receiving aerial horizontally along the roof on a line perpendicular to the direction of propagation. Such patterns must result from a secondary reflecting or scattering source flanking the direct path TR. In the present instance the "direct" path suffered a measured diffraction loss of 40 dB due to the severely obstructing shoulder of the Wrekin, thereby becoming very vulnerable to the influence of secondary paths.

The source of a single secondary path can be identified by noting that if the receiver R is moved horizontally at right angles to the dominant propagation path TR (Fig. 9), any offset source of multipath, S, will produce an interference pattern having a fringe distance given by

$$d = \pm \lambda / \sin \alpha \qquad \ldots\ldots(4)$$

where λ is the wavelength and α is the horizontal angle between the direct path and the path RS.

Knowing d and λ, the direction of the unknown source of multi-path can be determined from the relation

$$\alpha = \sin^{-1} [\pm \lambda / d] \qquad \ldots\ldots(5)$$

which is valid so long as $d \ll D \sin \alpha$, where D is the length of the direct path TR. This condition is well satisfied in the present case.

The mean values of d, measured from the two interference patterns (Fig. 8), yield $\pm 18{\cdot}1$ deg and $\pm 17{\cdot}9$ deg as the respective principal values of α for 173·4 and 183·4 Mc/s, leading to a mean value for α of $\pm 18{\cdot}0$ deg. The ambiguity of sign is easily resolved in this instance by the nature of the landscape flanking the direct path. A conjectural southward deviation of 18 deg at Ironbridge is ruled out because

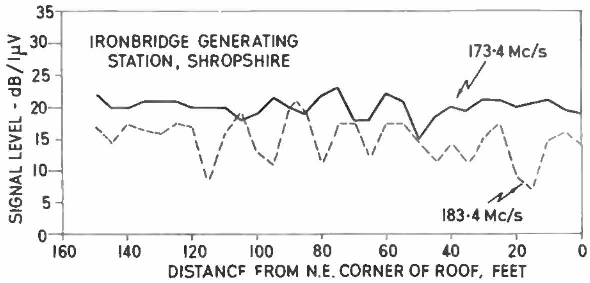


Fig. 8. Horizontal interference patterns measured at v.h.f.

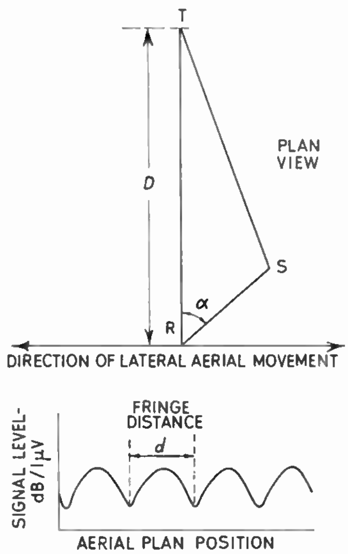


Fig. 9. Geometry of interference by lateral reflection.

the resulting line follows the shadowed Severn Valley, but an equal *northward* deviation gives a line intersecting the upper slopes of the Wrekin, as shown on Fig. 7. This is a plausible secondary path, because the narrow wedge-shaped area, S, which it intersects (Fig. 7) is visible from both terminals and therefore qualifies as a potential scattering or partially reflecting source. The curtailment of this area about half a mile south-west of the Wrekin summit is due to obstructing foothills in the Ironbridge direction, and it is for this reason that the secondary signal follows a route TSR that at first sight may have appeared somewhat arbitrary.

Scrutiny of the best-defined troughs of the interference patterns shows that the common scattering area contributes a secondary signal about 48 dB below the free space level, implying an effective "reflection coefficient" of about 0·004. The smallness of this coefficient is understandable when the magnitude of the scattering angle and the smallness of the common scattering area are considered.

The path under discussion introduces a secondary signal only 10 dB weaker than the direct one, and delayed on the latter by about 1 microsecond. While quite satisfactory for its specialized application, this path would of course be debarred from frequency-division-multiplex circuits on account of severe multipath distortion.

5. The Potentialities of Passive Repeaters

When intervening obstructions prevent direct microwave communication, a workable radio link is sometimes achievable by erecting a suitably angled flat mirror (or in certain situations a pair of mirrors) on a prominent site visible from each terminal. While the cost of installing and aligning the radio mirror system is inevitably high, its subsequent maintenance is negligible. The system is particularly attractive where the local topography provides a suitable elevated site within a mile or two of either terminal, in a situation where an active repeater would be costly to install and maintain. There is the further advantage of economy in the frequency spectrum.

5.1. *The Single-Mirror Passive Repeater*

Andrieux[10] has given basic design formulae for a system involving a single flat mirror (Fig. 10) situated far enough from the radio terminals T and R for the field contributions from all points on the mirror to be virtually in phase at the receiver. The latter restriction, which strictly speaking limits the mirror size to a substantial fraction of the first Fresnel zone, is never in practice violated to any serious degree.

Andrieux defines the efficiency of the system by the ratio

$$\frac{\text{power received via the mirror}}{\text{power received over a direct path of equal total length in free space}}$$

which amounts to

$$\frac{A_p^2(d_1+d_2)^2}{\lambda^2 d_1^2 d_2^2} \qquad \ldots\ldots(6)$$

where A_p is the mirror area projected on the aperture plane of the aerial T or R, d_1 and d_2 are as specified on Fig. 10 and λ is the wavelength.

For a practicable mirror size, good efficiency requires d_1 or d_2 to be small and the wavelength to be short, with moreover a large reflection angle θ (say 30 deg or greater) to ensure that the projected area A_p is not unduly diminished. Although unity efficiency, as defined by Andrieux, is not always essential, it is often desirable because it implies an overall performance no worse than would be obtained with the same total path length under free space conditions.

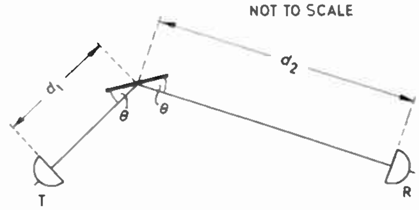


Fig. 10. Use of a flat mirror as a passive repeater.

For unity efficiency, the projected mirror area A_p must be

$$\frac{\lambda d_1 d_2}{d_1+d_2} \qquad \ldots\ldots(7)$$

and the necessity of limiting the mirror size usually requires in practice that $d_1 \ll d_2$ (adopting the convention that d_1 is the distance between the passive repeater and the nearer radio terminal). Thus for practical purposes the requisite projected mirror area for unity efficiency is

$$\lambda d_1 \text{ approx.} \qquad \ldots\ldots(8)$$

Figure 11 shows the approximate projected area required for unity efficiency at various wavelengths and mirror distances, d_1, from the nearer terminal. The actual mirror area must be

$$A_p \operatorname{cosec} \theta \qquad \ldots\ldots(9)$$

The requisite mirror size becomes impossibly large when θ is small, and the system becomes impracticable when θ is less than about 30 deg. Mechanical and economic considerations limit the mirror's projected area to about 1000 square feet. Floriani[11] furthermore advises restricting the *height* of the mirror to about 25 ft. On account of these limitations the use of radio mirrors is seldom practicable below 1500 Mc/s, as will be appreciated from Fig. 11.

The large dimensions of an efficient radio mirror involve a very narrow reflected beam, of the order of $\frac{1}{2}$ deg. Based on the idealization of a uniformly illuminated aperture, the reflected beam-width for 6 dB loss is approximately

$$70\lambda/a \text{ degrees} \qquad \ldots\ldots(10)$$

where a is the relevant mirror dimension and λ is the

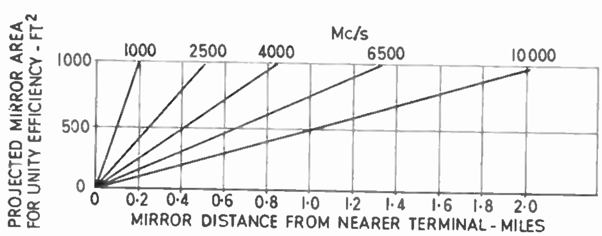


Fig. 11. Requisite projected mirror area for unity efficiency.

wavelength. While the narrow beam-width reduces fading and multi-path influences, it involves stringent mechanical requirements. The mirror must have close enough flatness tolerances to avoid significant diffuseness in reflection (Fig. 3), and means of extremely fine angular adjustment in azimuth and elevation. Furthermore the prescribed tolerances must be preserved in extremes of temperature and wind velocity. Floriani[11] prescribes flatness tolerances of $\pm\ \lambda/16$, and, in connection with the 6000 and 10 000 Mc/s bands, angular tolerances of $\pm$ 0·17 deg and $\pm$ 0·05 deg respectively for mirror dimensions of 10 ft and 30 ft.

5.2. *Double-Mirror Passive Repeaters*

When a passive repeater is collinear, or nearly so, with the radio terminals, the resulting small reflection angle at a single mirror gives a low reflecting efficiency. This can be restored by a double-mirror or "periscope" system, designed to secure a steep angle of ray incidence at each mirror. Besides the prerequisites of adequate size and steep reflection angles, double mirrors must not "shadow" each other, nor must the coupling loss between them be excessive. These requirements impose lower and upper limits on the spacing. Regarding maximum mirror spacing, it has been calculated that the coupling loss between two equal square apertures, one of which is uniformly illuminated, is about 2 dB when they are separated by half their Rayleigh range. This leads to a working rule for double-mirror systems, namely that the mirror spacing should not exceed half the Rayleigh range applicable to the shorter of the projected mirror dimensions.† The rule ensures that the coupling loss between mirrors never exceeds 2 dB and leads to permissible mirror spacings of about 250 ft in typical cases.

There are two types of double-mirror system, having differing properties worth a brief review.

5.2.1. The symmetrical double-mirror system

This system comprises two identical mirrors symmetrically arranged so that the incoming and outgoing rays are crossed. Figure 12 shows the limiting case when mutual shadowing is just avoided. Geometrical optics prescribe a reflection angle of $(90-\beta/4)$ deg at each mirror, where β is the angle of deviation between the incoming and outgoing ray paths.

To avoid shadowing of one mirror by another, the minimum permissible distance between mirror centres is given by

$$d_{\text{min.}} = \frac{l}{2}\operatorname{cosec}\frac{\beta}{4} \qquad \ldots\ldots(11)$$

where l is the length of each mirror and β has the previous connotation.

As β decreases $d_{\text{min.}}$ increases, until the maximum permissible spacing prescribed by the coupling loss between mirrors is reached. It can be shown that the above restriction limits the use of the system in the 4000 Mc/s band to situations where β exceeds about 10–15 deg. Regarding the upper limit of β, the system remains very efficient for values of β up to 90 deg, at which point it is however meeting severe competition from the single-mirror system. In this condition the latter has an efficiency only 2·4 dB

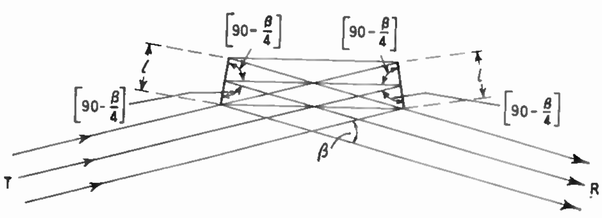


Fig. 12. Geometry of symmetrical double-mirror system.

worse than twin mirrors each of the same size. However for values of β less than 90 deg twin mirrors rapidly increase their advantage. To sum up, the symmetrical double-mirror system is very efficient when β lies between about 10 deg and 90 deg, but in view of its large total surface area, it may not be economically justifiable when β exceeds, say, 60 deg.

5.2.2. The unsymmetrical double-mirror system

To cater for the range of β between 0 deg and about 15 deg (i.e. the situation when the repeater is

† If a mirror has a dimension a_p after projection on the aperture plane of the illuminating aerial, the relevant Rayleigh range is defined as $a_p^2/2\lambda$ where λ is the wavelength.

nearly collinear with the terminals) an unsymmetrical double-mirror system must be used. This involves mirrors that are not parallel, having reflection angles differing by $\beta/2$, where, as illustrated on Fig. 13, β is once again the angle between incoming and outgoing rays. Figure 13 shows the system with the minimum possible mirror spacing, such that outer rays incident on mirror M1 just clear the edge of mirror M2, while at the same time allowing full illumination of the latter mirror.

To avoid shadowing of M1 by M2, the minimum permissible spacing between mirror centres is given approximately in this case by

$$d_{\text{min.}} = \frac{l}{2} \tan \left[\theta + \frac{\beta}{2}\right] \qquad \ldots\ldots(12)$$

where l is again the length of each mirror.

Study of eqn. (12) shows that the maximum possible value of θ, on which the system's reflecting

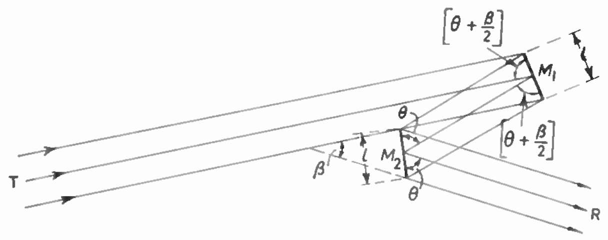


Fig. 13. Geometry of unsymmetrical double-mirror system.

efficiency basically depends, is limited firstly by the maximum mirror spacing prescribed by the coupling loss, and secondly by the angle β. The system has maximum efficiency when β is zero, thus admirably filling the gap where the previously described symmetrical system fails.

5.2.3. Comparison between the two systems

It is of interest to review the relative merits of the two systems. In the usable range of β the symmetrical system slightly excels the other in reflecting efficiency, but it fails, as explained, with small values of β. However, the unsymmetrical system gives adequate efficiency for β ranging between 0 and 60 deg. For example, an unsymmetrical system on 4000 Mc/s, having $\beta = 60°$ and mirrors 33 ft long by 16 ft high spaced 270 ft apart (the maximum permissible spacing for limiting the coupling loss to 2 dB), need have a performance only 1·3 dB worse than its symmetrical counterpart. This disparity becomes even smaller as β decreases. Finally, the unsymmetrical system has the advantage of a useful flexibility in design (denied to the symmetrical system) through choice of the reflection angle θ indicated on Fig. 13. For maximum efficiency θ must be made as large as is compatible with the maximum permissible mirror spacing. Sometimes however it is profitable to reduce θ somewhat, to allow a smaller mirror spacing on a restricted site. A substantial reduction in spacing can often be made at only small cost in reflecting efficiency.

Consider for example a restricted site allowing a mirror spacing of only 65 ft at 4000 Mc/s, β being 20 deg. Using the same size mirrors as in the previous example, a symmetrical system would need a spacing of 190 ft,† which is much too large for the restricted site. If instead, we design an *unsymmetrical* system to fit the maximum available space of 65 ft, its efficiency is found to be only 0·8 dB worse than the other system debarred on the grounds of excessively large spacing.

While the unsymmetrical system is satisfactory in any situation where a single mirror would be uneconomic, the symmetrical system forms a useful and slightly more efficient alternative in the somewhat restricted situations to which it is well adapted.

5.3. *Existing Mirror Installations*

Floriani[11] gives details of an unusual 6700 Mc/s single-mirror repeater on Monte Strabut in Italy, involving the exceptionally steep ray inclination of 68 deg by reason of the terminal at Tolmezzo, situated about 1 mile from the mirror in the valley immediately below. The distant terminal at Udine allows a reflection angle θ of 46 deg. This system, which requires an exceptional mirror tilt to accommodate the unusual path geometry, performed exactly as computed for a total path length of 27 miles. The system carries 120 voice channels with a mirror area of 320 ft^2. Floriani describes further a complex chain of mirrors and active repeaters comprising a 120-channel link connecting Milan, Como and Sondrio in N. Italy. The system operates in the 6000–7000 Mc/s band with mirrors of 320 ft^2, and is notable for the use in several instances of two successive mirror systems inserted between adjacent active repeaters. A typical link of this kind involves a low-lying active repeater near Lake Como, illuminating an adjacent mirror on the surrounding mountains, which deviates the beam towards a second mirror system some 25 miles away, which in turn deflects the beam down to the next active repeater by the lake side. This system, which lends itself well to the local topography, has the virtue of keeping all active repeaters in accessible low-lying regions. In all instances except one, where a symmetrical double-mirror system was used, the topography allowed of single-mirror repeaters. A proposed re-routing of this system will, according to Floriani, require mirrors of over 700 ft^2 and the installation of another double-mirror repeater.

† This distance is well within the prescribed maximum of half the appropriate Rayleigh range.

6. Some Influences of the Lower Atmosphere

It has been appreciated since about 1930 that wave propagation in the metric, decimetric and centimetric bands can be profoundly influenced by the troposphere. Spatial and temporal variations in the atmospheric refractive index amounting to a few parts in a million can promote large and sometimes rapid changes in received field strength.

In the frequency band under discussion, temporal field strength variations are controlled by the structure of the lower atmosphere, which in turn is greatly influenced by the land or sea surface below it. As shown by Booker,[12] the finer details of atmospheric stratification near the ground become progressively influential with shortening wavelength. Fortunately the advent of the cavity refractometer, in conjunction with conventional meteorological soundings, has enabled spatial and temporal changes in the atmospheric refractive index to be explored in sufficient detail to explain satisfactorily the main features of the radio phenomena observed.

6.1. *Ground-Reflection or K-Fading*

Where an optical path is subject to a near-specular ground or sea reflection, changes in the refractive index gradient of the atmosphere near the surface can produce large changes in the path difference between direct and surface-reflected waves, thereby causing deep interference fading. From the geometrical standpoint these atmospheric changes are equivalent to changes in the radius of a fictitious earth at which reflection via straight ray paths can be considered to take place. The often quoted effective earth radius factor, K, is the ratio of this fictitious earth radius to the true earth radius, R_0, and K is itself related to the refractive index gradient as specified at the beginning of this paper.

Owing to topographical and practical limits imposed on the heights of radio terminals above the intervening reflecting surface, and despite the fact that phase shifts increase with frequency, it is usually simple to ensure in the v.h.f. band that the path difference between direct and surface-reflected waves remains well below one wavelength, thereby excluding the possibility of fading by this mechanism. However on frequencies above about 500 Mc/s the phase shifts engendered by commonly occurring changes in refractive index gradient are often sufficient to include one or more states of destructive interference. Fading of this type is best considered as due to movement of a succession of interference fringes past the aerial as the refractive index gradient changes. For a specified change in gradient the number of fringes passed through, and hence the probability of fading, increases both with the frequency and the path length, providing always of course that the terminals remain high enough to keep in the interference region. In the upper v.h.f. and lower u.h.f. ranges the adverse effect of surface-reflections may sometimes be minimized by adjusting the height of either terminal for maximum signal strength under "standard" atmospheric conditions. This expedient is however quite useless when atmospheric changes cause a shift of many fringes. It is a simple matter when planning optical paths to calculate the movement of fringes likely to occur through atmospheric changes. A simple working rule is to assume an effective earth radius factor ranging between 1 and infinity, although this does not cater for extreme sub-refracting conditions when K may fall as low as 0·8.

Interference fading of the type described is severe only when the surface reflection coefficient approaches unity. The employment of vertical polarization with its characteristically lower surface reflection coefficient can substantially reduce the fading over water paths as earlier explained. In fact 1000 Mc/s tests by Quarta[2] over an optical sea path showed fades of about 15 dB with vertical polarization, whereas the fading increased to about 25 dB using horizontal polarization.

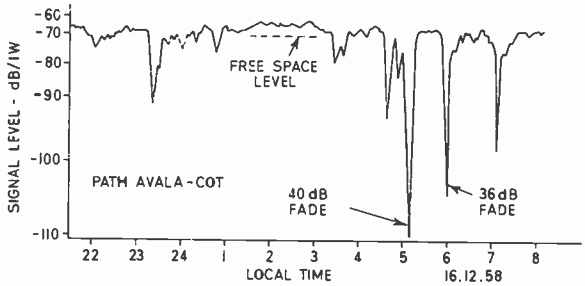


Fig. 14. Ground reflection fading, Yugoslavia, 2000 Mc/s.

Fading of the present type can also occur over land, but fortunately less frequently than over water because of diffuseness in the ground reflection due to ground roughness. Figure 14 illustrates such a case of fading with horizontal polarization over a 2000 Mc/s path in Yugoslavia, involving high terminals above a broad and flat river valley. Owing to the high Brewster angle for land (about 20 deg), coupled with the grazing reflection angle (about 1 deg), it is doubtful whether the use of vertical polarization would significantly reduce the fading over this path.

Several other observers have recorded severe K-fading over optical paths ranging in length between about 40 and 70 miles. Bray and Corke[13] have measured fades as deep as 40 dB in 3930 Mc/s tests across the Irish Sea. Bogle,[14] using 3200 Mc/s, and Cabessa[15] using 2000 Mc/s, have observed maximum fades of about 35 dB.

6.2. *Properties of Reflecting Atmospheric Layers*

As is well known, meteorological processes often promote horizontal stratification of the atmosphere, resulting in rapid changes in the gradient of atmospheric refractive index. Frequently a temperature inversion, combined with a lapse in humidity, conspires to give a very steep gradient of refractive index over a narrow height interval. When such a layer lies above the radio terminals, as commonly happens over flat or undulating ground cooling by radiation on a clear night, a substantial reflection at grazing incidence to the layer can produce fading by interference with the direct signal. This influence is the dominant cause of fading in the v.h.f. band, and a contributory agent at higher frequencies, occurring mainly during night and early morning over land paths in response to the occurrence of elevated nocturnal inversion layers. It is proposed briefly to review the properties of such layers, with particular reference to the behaviour of a non-optical v.h.f. path tested in the Arabian Gulf, where reflections from recurrent nocturnal inversion layers habitually swamped the weak diffracted signal received in their absence.

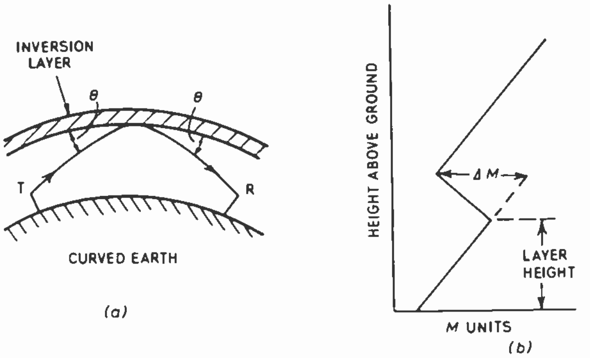


Fig. 15. (*a*) Ray geometry for reflection at an inversion layer. (*b*) Specification of *M*-deficit, ΔM.

6.2.1. Theoretical considerations

Millington[16] has evolved curves for computing the reflection coefficient of horizontally stratified layers, using an idealization involving a suitably graded layer with zero refractive index gradient outside it. In practice the refractive index gradient is usually negative in those surrounding regions, and this departure from the prescribed idealization reduces the grazing ray angle on account of downward ray bending as indicated diagrammatically on Fig. 15 (*a*). To allow for this effect, while at the same time conforming to Millington's results for layers suiting his idealization, the present author has devised a simple modification to Millington's fundamental parameter involving the total change in refractive index across the layer. This empirical adjustment involves an *M*-deficit, ΔM, associated with the modified refractive index profile for the layer, as geometrically specified on Fig. 15 (*b*). Millington's change in refractive index is then interpreted as

$$\Delta M \times 10^{-6} \qquad \ldots\ldots(13)$$

where ΔM is expressed in *M* units.

For layers having a negative gradient of refractive index and finite thickness, Millington has shown that total internal reflection occurs when the grazing angle is less than a critical value dependent only on the total change of refractive index across the layer (effectively the modified parameter ΔM just defined). For larger grazing angles the reflection coefficient declines in a rapid but complex way governed by the layer thickness expressed in wavelengths.

The critical angle for total internal reflection is given by

$$\theta_c = 10^{-3}\sqrt{2\Delta M} \quad \text{radians} \qquad \ldots\ldots(14)$$

where ΔM has the previous connotation. Meteorological upper air soundings have shown that ΔM seldom exceeds 100 *M* units, so that θ_c is limited to about 0·8 degree.

The occurrence of total internal reflection at sufficiently grazing incidence explains the "skip distance" effects often seen on radars during "anaprop" conditions.[17] A target just beyond the radar's normal range may involve too steep a reflection angle at an elevated layer to reflect sufficiently to give an anomalous return. However a more distant target may well involve a grazing angle shallow enough for total internal reflection, thereby providing effective "intervisibility".

6.3. *A V.H.F. Test in the Arabian Gulf*

Further evidence of the predicted occurrence of total internal reflection has accrued from prolonged tests on 80 Mc/s over a non-optical 85 mile path in the Arabian Gulf,[18] comprising part water but mostly land. Meteorological upper air soundings made by the R.A.F. at Bahrain have revealed the periodic occurrence of intense nocturnal inversion layers. Modified refractive index profiles of six of these elevated layers are shown on Fig. 16. All these soundings, made at 0600 hours local time and selected from a test period of 5 months, were associated with the simultaneous reception of abnormally strong signals on 80 Mc/s over the neighbouring radio path between Bahrain and Doha, which was tested continuously for the above period. Application of Millington's process for computing reflection coefficients of elevated layers shows that the illustrated *M*-profiles can all be expected to give total internal reflection over the tested radio path. This is consistent with the occurrence of signal levels near the free space value that in every case were

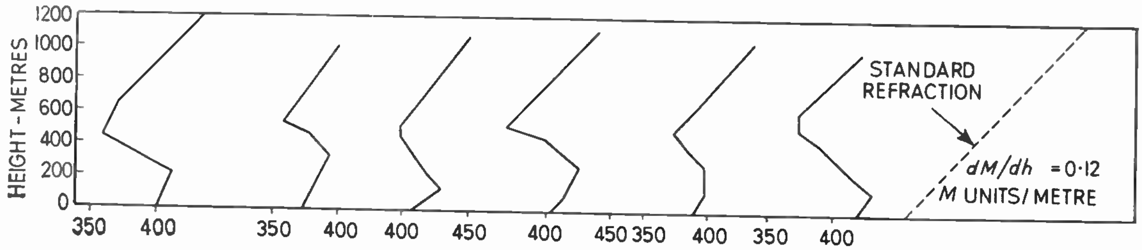


Fig. 16. Modified refractive index profiles showing elevated inversion layers at Bahrain (Synoptic hour, 0600 local time).

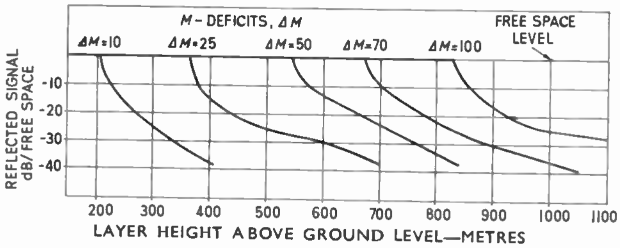


Fig. 17. Theoretical relations between layer-reflected signal strength and layer height for 80 Mc/s Bahrain–Doha path (Layer thickness 300 m).

associated with these selected soundings. It should be emphasized that under the fairly "standard" atmospheric conditions habitually occurring round midday, the test path gave a signal strength about 50 dB *below* the free space level, thus providing a very sensitive index of anomalous atmospheric conditions.

Based on the principles already described, Fig. 17 shows theoretical signal strengths on 80 Mc/s expected from reflection at hypothetical layers 300 m thick at various heights above the radio path in question. To conform broadly to the *M*-profiles under discussion, a representative range of *M*-deficits is exhibited and the curves are based on a uniformly graded atmosphere below the layer corresponding to an effective earth radius factor of 2. Signal strengths are expressed relative to free space. Although the curves should not be taken too literally, they serve to illustrate the very high signals resulting from reflection at even minor atmospheric irregularities if the ray incidence is sufficiently grazing.

Figure 18 shows examples of 80 Mc/s signal strength recordings over the Bahrain-Doha path. Figure 18 (*a*) shows nocturnal fading from a high level resulting probably from interference between strong multiple

(*a*) Fading due to nocturnal inversion layer.

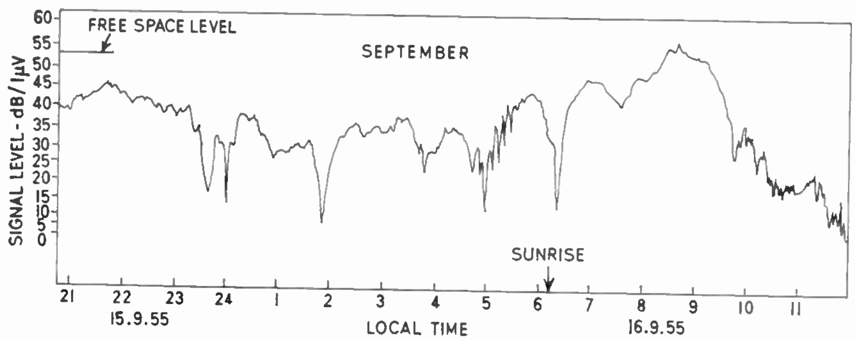
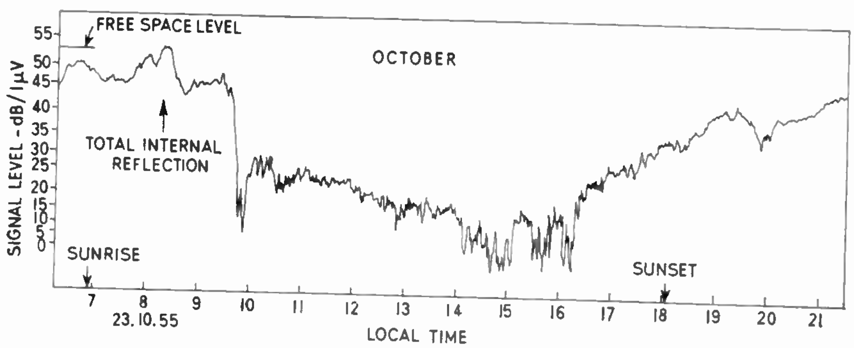


Fig. 18.

Illustrative signal strength recordings on 80 Mc/s over Bahrain–Doha path.

(*b*) Representative diurnal behaviour.



reflections at various points on an elevated inversion layer. Figure 18 (*b*) shows a persistent high level during early morning, attributable to relatively undisturbed total internal reflection at the layer. A striking feature is the rapid drop of 40 dB in signal strength at about 0930, indicating the sudden dispersal of the layer by morning convection. A slow return of atmospheric stratification after 1700 is also evident.

6.4. *Drop-outs due to Total Internal Reflection*

A sometimes-overlooked consequence of total internal reflection is the failure of a totally reflected transmission to be receivable in certain regions above the layer. This can make a radar "blind" to aircraft unfavourably situated above an inversion layer, and can occasionally cause drop-outs of long duration in point-to-point links, where a low-lying inversion layer can intervene between the respective radio terminals. Fortunately this condition is rare because, as shown by eqn. (14), the ray incidence must be very shallow for the effect to occur.

6.5. *Propagation in Ducts*

The basic features of propagation within a duct have been simply explained by Appleton,[19] by considering interference between upgoing ground-reflected waves and downward-refracted waves within an M-inversion. This simple treatment shows quite satisfactorily the vital role played by the wavelength in duct propagation, and it is in fair numerical agreement with conclusions based on Booker's[12] development of the mode theory of propagation. Without entering deeply into the numerical consequences of the mode theory it may be stated that when the terminals lie within an M-inversion extending over the track-widths of the first or higher-order propagation modes, signal strengths within the duct may be expected to be very high at points well beyond the transmitter's geometrical horizon. The non-optical Bahrain-Doha path, already discussed, has amply confirmed this expectation at 80 Mc/s, as will be shown.

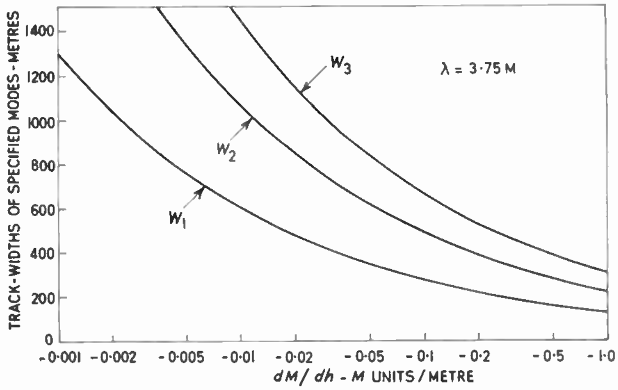
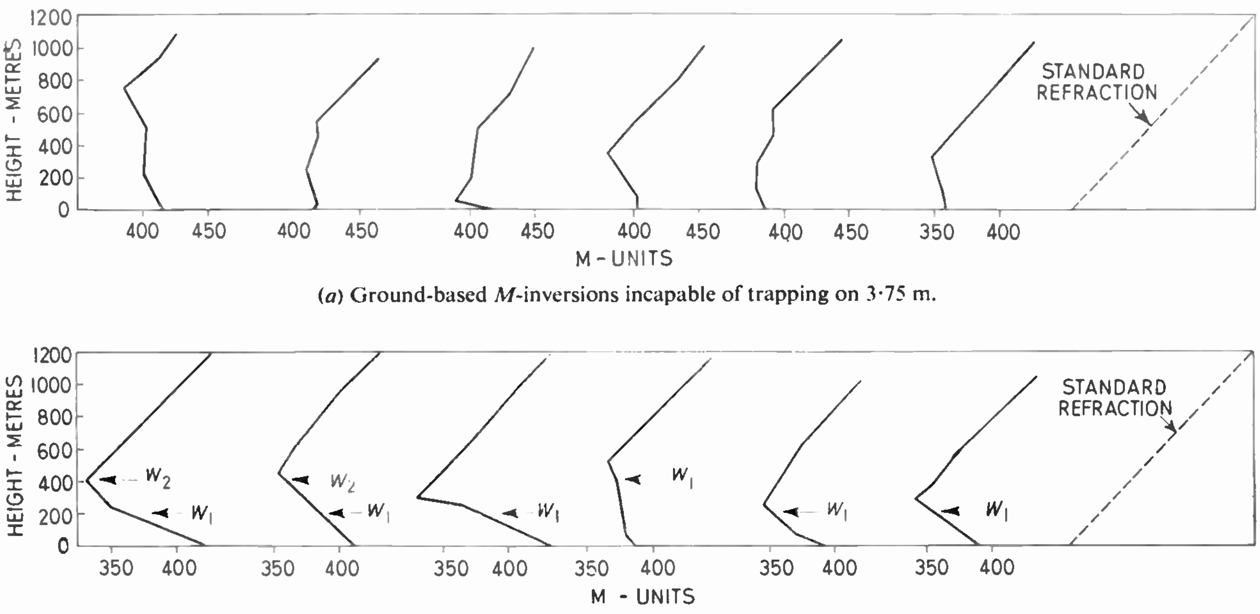


Fig. 19. Theoretical relations between dM/dh and respective track widths for the first three propagation modes on a wavelength of 3·75 m. (After Booker.[12])

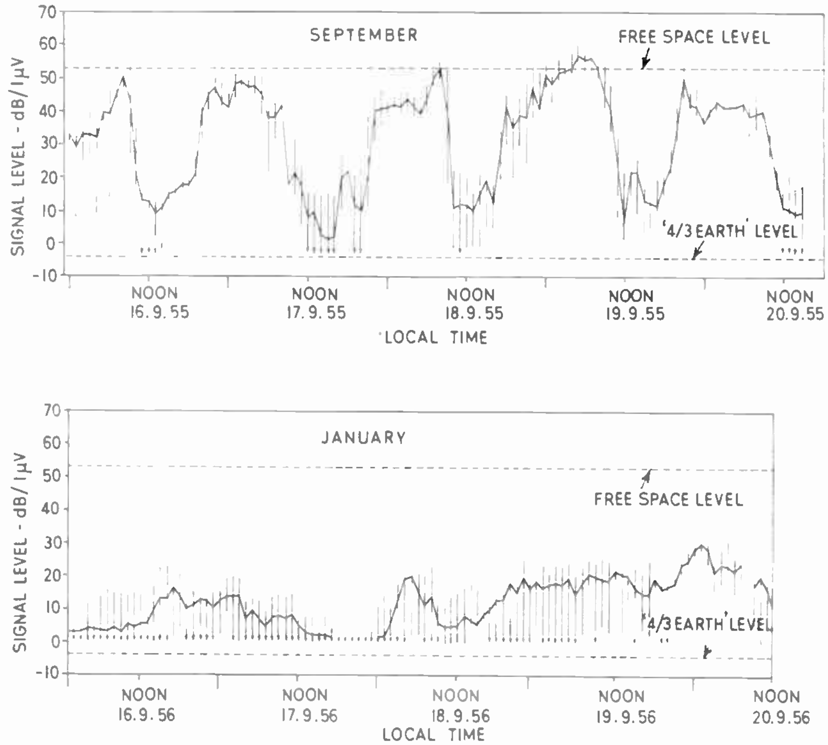
The characteristic propagation modes are associated with track-widths whose size increases with the wavelength and the order of the mode. We are concerned with negative values of modified refractive index gradient—the duct condition—where the track-width is precisely defined. Based on Booker's work,[12] Fig. 19 shows the relation between the modified



(*a*) Ground-based M-inversions incapable of trapping on 3·75 m.

(*b*) Ground-based M-inversions capable of trapping one or two modes on 3·75 m.

Fig. 20. Modified refractive index profiles showing ground-based inversion layers at Bahrain (Synoptic hour, 0600 local time).



(*a*) Diurnal variations.

Fig. 21.
Hourly signal ranges, 80 Mc/s, Bahrain–Doha.

(*b*) Random variations.

refractive index gradient, d*M*/d*h*, and the track-widths of the first three propagation modes for the test frequency of 80 Mc/s (3·75 m).

During the Bahrain-Doha tests, upper air soundings at Bahrain showing ground-based *M*-inversions were found associated with high signal levels on 19 occasions. A selection of the *M*-profiles, all obtained at 0600, local time, is given in Figs. 20 (*a*) and 20 (*b*). To determine the number of transmission modes trapped by a particular duct, the track-widths of the first three modes were derived from Fig. 19, using an average value of d*M*/d*h* near the surface obtained from the relevant *M*-profile. In seven cases, of which six are illustrated on Fig. 20 (*a*), the track-width of the first mode significantly exceeded the height of the *M*-inversion, indicating absence of trapping. In the remaining 12 cases the *M*-inversion covered the track-width of the first mode, while in two of these cases the second mode was just trapped also. Figure 20 (*b*) shows examples of this last category. Where a mode is trapped the top of the associated track is indicated with an appropriately marked arrow.

There is ample documentary evidence of trapping by maritime ducts on wavelengths as long as 1½ m, but observations of trapping on land paths are more scanty. There is however war-time evidence from 200 Mc/s radar observations in Bengal.[20] When looking westwards over the flat Ganges delta, these radars experienced anomalous ranges of up to 200 miles on most nights during the season October–May, whereas their "orthodox" day-time range was only about 40 miles. Local meteorological soundings clearly point to nocturnal surface ducts as the agency during the dry season October–February, while the continuance of "anaprop" into the hot season March–May is attributable mainly to total internal reflection at elevated wet ducts.

On wavelengths longer than 1½ m, the literature provides little well-substantiated evidence of trapping as rigorously defined by Booker.[12] Heightman[21] has made v.h.f. tests over non-optical paths terminating at a coastal radio station at Clacton. His oversea tests on 145 Mc/s with fixed stations on the continent and with a ship in the North Sea yielded fine-weather signal levels sometimes approaching the free space value at distances well beyond the horizon of the Clacton terminal. Thanks to the low elevation of the stations and the possibility of pronounced maritime ducting, trapping may have occurred in some of these instances, but this is conjectural in the absence of accompanying meteorological soundings and in view of alternative mechanisms of anomalous propagation that might have been operating. Similar behaviour has been noted by Heightman[21] at 59 Mc/s on a 120 mile over-land path in the United Kingdom, but in the present author's opinion no trapping occurred there, because it is most unlikely that land in temperate regions could experience a surface duct having the M-gradient and thickness required for trapping on this relatively low frequency.

On the contention that claims to trapping at low frequencies require substantiation by appropriate meteorological measurements, it is submitted that the previously described Bahrain-Doha tests on 3·75 m confirm the existence of trapping on a longer wavelength than has perhaps been authenticated hitherto.

6.6. *Diurnal and Seasonal Variations*

To illustrate the profound influence of diurnal and seasonal atmospheric changes on long v.h.f. radio paths suffering extremes of climate, some statistics will be presented from the prolonged test on 80 Mc/s over the Bahrain-Doha path[18] already discussed. Signal strengths over this 85 mile path were continuously recorded from August 1955 until March 1956.

Figure 21 shows charts of hourly signal ranges, on a scale of local time, typical respectively of September and January. During September the difference between mean daytime and night-time levels often amounted to 40 dB and diurnal changes were repeated with remarkable regularity in that month. During January diurnal patterns were only occasionally perceptible. Figure 18 (*b*), previously discussed, shows a typical record of diurnal behaviour. It exhibits,

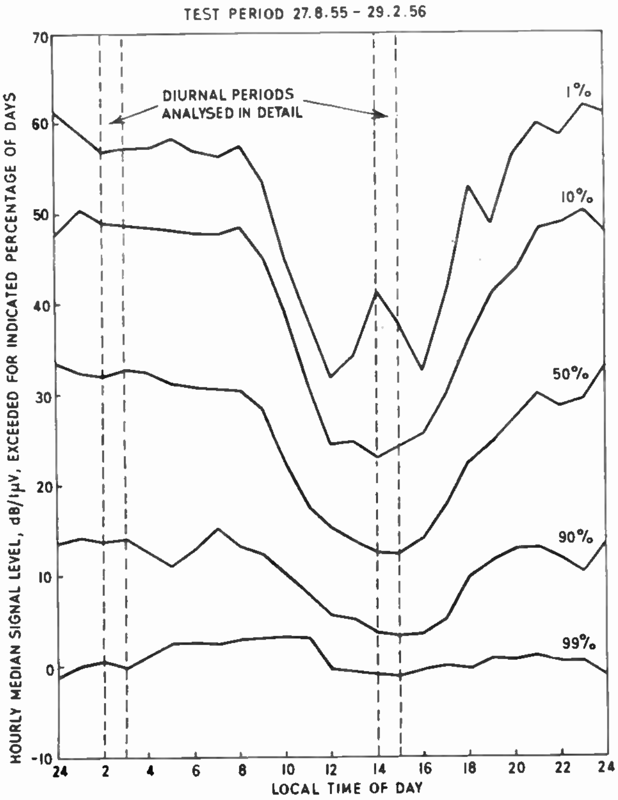


Fig. 22. Statistics of diurnal variations, 80 Mc/s, Bahrain–Doha.

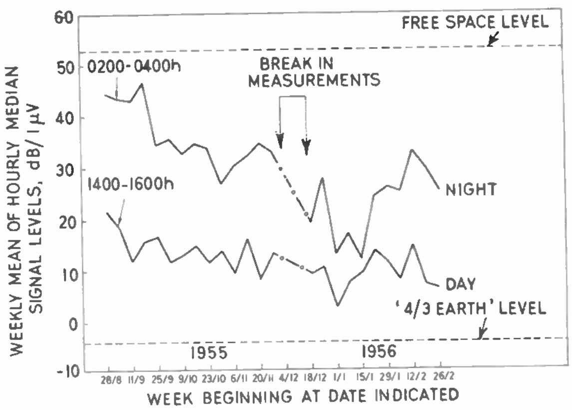


Fig. 23. Seasonal trends for daytime and night-time periods, 80 Mc/s, Bahrain–Doha.

firstly, an early morning period when the strong received signal is derived from a near-specular reflection at an elevated inversion layer. Secondly, there is an early afternoon period consistent with a near-standard atmosphere and thirdly, there is a steady growth in signal strength after 1600 hours indicative of the returning nocturnal inversion.

Figure 22 summarizes the statistical behaviour of the diurnal variations over the 6 months test period. Each curve shows for the specified times of day, the hourly median level exceeded for the indicated percentage of the 160 complete days tested. It is seen that the weaker hourly median levels, namely those exceeded during a high percentage of days, show progressively reduced diurnal variations, until by the time the 99% level is reached the diurnal effect is barely perceptible, whence it is inferred that the troughs of fades attain a fairly constant low level which is independent of the time of day. Attention will now be briefly directed to the behaviour during the period 1400–1600 hours when the signal strength is usually least, and the contrasting period 0200–0400 hours when the hourly median signal strength is near its maximum. Figure 23 shows the seasonal trends for these extreme periods of the diurnal cycle. The hourly median levels for the respective day and night periods have been averaged over each week of the test. The seasonal trend, with a minimum in January, is seen to be far more pronounced for the night period than for the day period. Extrapolating to the unrecorded seasons, it seems likely that both daytime and night-time weekly mean levels would show broad maxima in early July.

As is now well known, the main agent in diurnal disturbances is the nocturnal cooling of heated land by radiation when the sky is clear, which gives rise during the night and early morning to intense atmospheric stratification near the ground. Diurnal changes are generally more pronounced in summer than in winter, because higher temperatures increase the humidity lapse rate, thereby increasing the refractive index gradient near the surface. For this reason diurnal effects tend to be more marked in tropical regions than in temperate zones. For sea paths, particularly when they are well clear of coast lines, such variations may be intermittent or totally absent, for reasons connected with the sea's thermal stability.

6.7. *Effect of Random Atmospheric Irregularities*

With shortening wavelength, propagation is increasingly influenced by random irregularities in the atmospheric refractive index, whose existence has been directly revealed in fine detail by microwave refractometer measurements made by Essen[22] in the laboratory and by Bussey and Birnbaum[23] in aircraft. By using 4000 Mc/s pulses of 3 millimicroseconds duration over a 28 mile path devoid of ground reflections, de Lange[24] has revealed the simultaneous presence on summer nights of two or three distinct radio paths through the atmosphere, resulting from the irregularities just discussed. Pulses travelling by these paths had a maximum time separation of 7 mμs, corresponding to a 7-foot path difference. It was inferred from fading observations that path differences of only 6 inches were sometimes present. In this situation a c.w. transmission would be subject to interference-fading between the various paths.

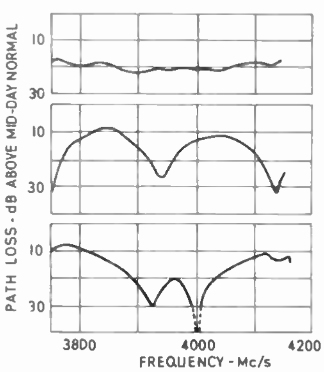


Fig. 24. Typical path loss-frequency curves. (After Kaylor.[25])

Several other workers have strikingly illustrated this effect with oscillograph displays of changing microwave signal levels as the transmitted frequency was quickly and linearly varied through 400 Mc/s in the 4000 Mc/s band. Figure 24 shows examples of traces obtained by Kaylor[25] over a path in Iowa having negligible ground reflections. During summer nights the traces often changed significantly within a few seconds. Mathematical synthesis of the observed waveforms revealed the amplitudes and delay times of the significant signal components.

Crawford and Jakes[26] have measured vertical angles of arrival on 24 000 Mc/s by nodding an aerial of 0·12 deg vertical beam-width through an angle of 2 deg. The system could detect changes of 0·05 deg in arrival angle. During daytime a single ray was usually indicated, while at times of exceptionally strong signal this widened into a solid wedge of rays covering a sector of 0·8 deg. At night three individual rays separated by about 0·25 deg were sometimes seen, or two rays about 0·5 deg apart appeared.

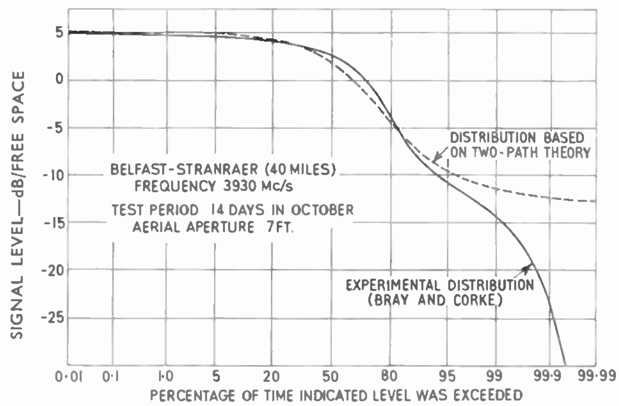


Fig. 25. Cumulative distribution for oversea path. (After Bray and Corke.[13])

The three techniques just described convincingly demonstrate that, under temperate summer conditions, microwave paths may carry three or more signal components with maximum delays of about 10 mμs. De Lange notes that this dispersion limits the ultimate performance of microwave multiplex systems.

6.8. *Some Implications of Fading Statistics*

Random processes inherent in microwave propagation through an inhomogeneous atmosphere necessitate a statistical approach to many problems. Indeed, statistical examination of signal strength recordings often reveals or confirms the nature of the propagation mechanism at work.

The full line curve of Fig. 25 shows a cumulative distribution of signal levels due to Bray and Corke,[13] derived from a fortnight's measurement on 3930 Mc/s of a 40-mile optical path across the Irish Sea employing horizontal polarization. This path, as already briefly mentioned, suffered severe fading due to adversely phased reflections at the sea surface. It is revealing to compare the measured distribution with a theoretical one (dotted curve in Fig. 25) based on the justifiable assumption of randomly phased interference between direct and sea-reflected signals resulting from fluctuations in the effective earth radius. For lack of more topical information, the author has assumed the variations in the effective earth radius to be statistically similar to those inferred from prolonged v.h.f. tests made in 1951 over a sea path in the Mediterranean.[27]

Agreement between the theoretical and measured distributions is good for levels exceeded for less than 95% of the time. The progressive divergence above 95% is associated with the unexpectedly deep fading often noted by investigators of this mechanism. The discrepancy with theoretical behaviour can be explained by the presence of a third small randomly phased component (itself possibly the resultant of many others) caused for example by turbulence or weak reflection at an inversion layer.

When a signal comprises many components having various amplitudes and random phases (as in transmission through an inhomogeneous atmosphere), the resultant signal levels have a Rayleigh distribution so long as no slow changes in mean level occur during the test period. This type of distribution, indicative of multiple ray paths, is evident from long-term microwave measurements on sea paths made by Cabessa[15] and by Gudmandsen and Larsen.[28] Sometimes however these paths faded more severely than compatible with Rayleigh's law, suggesting reversion to the idealized two-path propagation dominant in Bray and Corke's tests. In general, sufficiently short microwave sea paths suffer nearly pure two-path interference, but with increasing path length multiple ray paths in the troposphere become progressively significant until Rayleigh conditions are dominant. It should be emphasized that two-path fading (Fig. 25) can be substantially worse than the Rayleigh fading† to which long radio paths are often subject.

The statistical investigation of fading mechanisms is complicated by field variations across the aerial due to multi-path effects. Increasing the aerial size increases the chance, at any instant, of a usable field somewhere in the aperture, thereby reducing the chances of fading in some circumstances. From the transmission viewpoint, the narrow beam-width associated with a larger aerial excludes indirect ray paths that would otherwise contribute interference effects. However, as long as the beam is still wide enough to include at least two ray paths the chance of deep fading remains. Enough has been said to show that under multi-path conditions the fading characteristics of the received signal must depend in a complex way on the size of aerial employed. The simplest case, that of a large aerial subjected to two-path propagation, is discussed in the next section.

† Under Rayleigh conditions, fading exceeds 29 dB relative to the median level for 0·1% of the time.

6.8.1. Aperture integration

When an interference field set up by multiple ray paths has a scale of irregularities comparable with the size of a large receiving aerial, the received signal may fluctuate less than if the aerial were smaller, just as a star viewed through a large telescope twinkles less than when viewed by the naked eye. It is possible to treat the simple case of aperture integration by a horn or paraboloid aerial in the presence of sinusoidal vertical interference fringes resulting from specular surface reflection of the kind already discussed. The interwoven curves of Figs. 26 (*a*) and (*b*) show the extremes of signal pick-up (and thus the fading range) experienced by the specified large-aperture aerials as they are scanned by a complete interference fringe of height f, in response, as previously explained, to a change in the effective earth radius. The fluctuation in aerial pick-up resulting from the movement of a complete fringe past the aerial depends on the aperture-fringe ratio a/f (Fig. 26). The magnitude of this

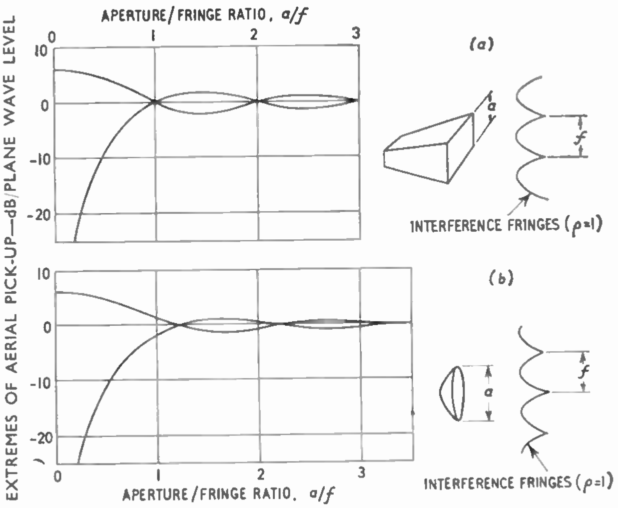


Fig. 26. Behaviour of wide aperture aerials subject to interference fringes.
(*a*) Rectangular aperture uniformly illuminated.
(*b*) Circular aperture uniformly illuminated.

fluctuation (i.e. the fading range) is given in decibels from the chart by the vertical interval between the limiting curves at the appropriate value of a/f. The charts are based on sinusoidal interference fringes (implying a surface reflection coefficient of unity) and uniformly illuminated apertures. Tapering the illumination alters the shape of the curves slightly, but the simple assumption of uniform illumination adequately serves to illustrate the principles involved.

The charts show how an increasing aperture-fringe ratio reduces the peak-to-trough range, until when this ratio reaches unity the fading becomes negligible. Furthermore, for ratios exceeding unity the aerial collects virtually the same signal as would be obtained in the absence of the surface-reflected wave. Mathematical integration shows that the effect of a large aerial is equivalent to reducing the reflection coefficient of the surface promoting the interference fringes.

For the rectangular horn, the reflection coefficient becomes effectively

$$\left| \frac{f}{\pi a} \sin \left[\frac{\pi a}{f}\right] \right| \qquad \ldots\ldots(15)$$

For the paraboloid, the reflection coefficient is effectively

$$\left| \frac{f}{\pi a} 2 J_1 \left[\frac{\pi a}{f}\right] \right| \qquad \ldots\ldots(16)$$

where J_1 is a Bessel function of the first order.

Considering a large aerial from the transmission viewpoint, it is easily appreciated that increasing its size eventually narrows the beam sufficiently to avoid illuminating the reflecting surface (e.g. the sea), thereby eliminating the interference fading.

While the limitations of aerial size and path geometry often make it impossible to achieve an aperture-fringe ratio near unity, the present curves show that a ratio as low as $\frac{1}{4}$ confers a useful benefit when coherent two-path propagation predominates. Measurements by Cabessa[15] over a 56-mile sea path confirm this conclusion. Using small aerials of unspecified aperture on 1200 Mc/s, Cabessa measured peak-to-trough fades as high as 37 dB, while simultaneous measurements on 2000 Mc/s with 10 ft diameter paraboloids showed fades of only 25 dB. The present author computes that for the latter test an aperture-fringe ratio of about 1/3 operated at the controlling terminal, giving a theoretical peak-to-trough fading range of 22 dB on the assumption of two-path propagation. The slight excess of the measured over the theoretical fading suggests the presence of subsidiary tropospheric ray paths, which is in keeping with the implications of Cabessa's long-term measurements already discussed.

6.9. *Microwave Diversity Systems*

As is well known, fades due to multi-path transmission are not always synchronized when separately received at two points spaced a sufficient distance apart. A like effect occurs when two transmissions over a common radio path differ sufficiently in frequency. In the frequency band under discussion, space and frequency diversity systems exploiting this principle can provide a useful increase in circuit reliability. The subject has been treated theoretically by Jelonek, Fitch and Chalk,[29] with illustrative statistics derived from microwave measurements made in Wales. Results of microwave space diversity tests over long optical sea paths have been given by Cabessa,[15] and factors governing the optimum diversity aerial spacing for over-water microwave paths have been investigated by Lewin.[30]

For maximum diversity advantage, fading on one channel must always be accompanied by high signal strength on the other, such that the correlation coefficient between the respective signals is -1. This ideal condition, termed by Bateman[31] "complementary diversity", can only exist where interference occurs between no more than two ray paths, as for example on short optical paths over sea (Section 6.1). Often, however, atmospheric irregularities introduce multiple ray paths as already discussed. In this situation an appropriate double-diversity system yields a choice of Rayleigh-distributed signals that ideally are independent (correlation coefficient zero), but in practice are often positively correlated. However even when positive correlation coefficients as high as 0·5 occur, useful diversity improvements are still obtainable, as shown by Grisdale, Morris and Palmer.[32]

6.9.1. Complementary diversity

With short optical microwave paths over sea or flat ground, two-path interference by surface reflection causes an approximately sinusoidal variation of field strength with height (Section 2.3), giving a vertical interference pattern as illustrated on Fig. 27 (*a*). Near R, the height interval between a point of virtually zero signal and the nearest point of maximum signal is given by

$$S = \lambda D/4H \qquad \ldots\ldots(17)$$

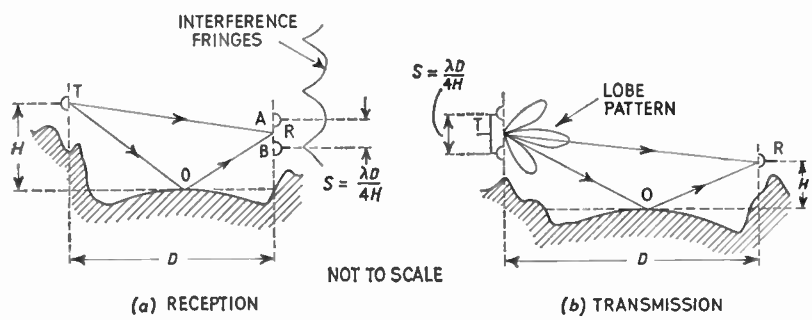


Fig. 27.
Complementary diversity system. (After Bateman.[31])

where λ, D and H are respectively the wavelength, path length and the transmitter height above the tangent to the reflection point at O. Bateman[31] has devised a "complementary diversity" system whereby two receiving aerials A and B are vertically spaced by the specified amount, S, as shown in Fig. 27 (*a*). When the effective earth radius alters, the interference fringes move bodily past the aerial system with little change in the value of S for paths well within visual range. Thus zero signal from one aerial is always accompanied by a large signal from the other. Bateman suggests paralleling the aerials to secure the benefits of diversity reception with a single receiver. It can be shown furthermore that, if the resulting aerial array is used for transmitting (Fig. 27 (*b*)), it duly directs a null towards the reflecting sea at O and a lobe towards the receiver at R, thus virtually eliminating the surface reflection and once again suppressing the associated fading.

6.9.2. Conventional space diversity

Conventional space diversity involves selecting the stronger of the respective signals from two aerials sufficiently separated to give zero or small positive correlation between them. Brennan[33] states that a diversity spacing of 30–50 wavelengths (which conforms to that adopted by other workers) is typically required for a correlation coefficient of 0·3.

Long optical sea paths involve a mixture of coherent and incoherent signal components, and under these circumstances Cabessa[15] recommends a diversity aerial spacing that suppresses the interference between direct and sea-reflected paths (namely the coherent interference). Such a system, which requires an aerial spacing given by equation (17), behaves almost as if the reflecting sea were absent, and hence has to contend only with the incoherent tropospheric ray paths.

Adoption of the above rule has given good results with 56 and 72-mile sea paths in the Mediterranean, over which Cabessa has made prolonged tests on 2000 Mc/s. In each case aerials spaced about 20 ft vertically showed a long-term diversity behaviour closely approximating to that theoretically expected from selection of the stronger of two independent Rayleigh-fading signals. Although the resulting "Rayleigh-diversity" law implies zero correlation between the two signals on a long-term basis, negative correlation associated with reduced fading sometimes occurred due to the occasional dominance of two-path propagation. Gudmandsen and Larsen[28] have noted similar diversity behaviour on microwave sea paths off the coast of Denmark.

Figure 28 shows a representative diversity record obtained from Cabessa's 72-mile sea path, which he considers to involve about the limiting distance for a

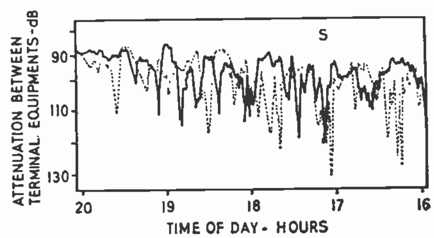


Fig. 28. Space diversity reception over a sea path. (After Cabessa.[15])
Path length, 75 miles.
Frequency, 2000 Mc/s.
Aerial spacing, 20 ft.

high-quality microwave link. While major fades on the respective channels are mostly out of step, there is occasional synchronism (at S for example), which statistically is to be expected. Whereas fading on either channel attains 30 dB at times, the long-term diversity fading is limited for 99·9% of the time to about 16 dB below the median level (the theoretical "Rayleigh-diversity" figure), which is a substantial improvement on the Rayleigh fading of 29 dB obtained with a single aerial. In fact, taking due account of a small benefit to the median level, diversity working has improved the path reliability by 15 dB on a 99·9% basis.

In conclusion, it seems that the use of space or frequency diversity systems designed to suppress interference from surface reflections will in practice confine microwave fading to "Rayleigh diversity" behaviour. This performance is likely to be the best that is consistently obtainable with double-diversity systems on long microwave links, whether oversea or overland. The diversity systems just reviewed are certain to find increasing application in extending the length of point-to-point links, while still maintaining a satisfactory standard of reliability.

7. Acknowledgments

The author is indebted to the Director of Research, Marconi's Wireless Telegraph Company Limited, for permission to publish this paper. He is also indebted to the Midlands Electricity Board for permission to reproduce Figs. 7 and 8, to the Directorate of Norwegian Telecommunications and the Board of Swedish Telecommunications for permission to publish Fig. 6, and to the Department of Posts and Telecommunications, Ghana, for permission to publish Fig. 4.

The author further acknowledges the work of Dr. D. H. Shinn in connection with the coupling loss between radio mirrors (Section 5.2), and the assistance of Mr. F. Immirzi with calculations relating to Figs. 25 and 26.

8. References

1. D. E. Kerr, "Propagation of very short waves—Pt. 2", *Electronics*, **21**, p. 118, February 1948.
2. P. Quarta, "Propagation with horizontal and vertical polarization over the Portofino–Monte Beigua and the Portofino–Monte S. Nicolao paths", *Alta Frequenza*, **26**, No. 5, p. 404, October 1957.
3. M. P. Bachynski, "Microwave propagation over rough surfaces", *R.C.A. Review*, **20**, No. 2, p. 308, June 1959.
4. L. H. Ford and R. Oliver, "An experimental investigation of the reflection and absorption of radiation of 9-cm wavelength", *Proc. Phys. Soc. (Lond.)*, **58**, p. 265, May 1947.
5. J. P. Day and L. G. Trolese, "Propagation of short radio waves over desert terrain", *Proc. Inst. Radio Engrs*, **38**, p. 165, February 1950.
6. G. W. G. Court, "Determination of the reflection coefficient of the sea for radar coverage calculation by an optical analogy method", *Proc. Instn Elect. Engrs*, **102**, Pt. B, p. 827, November 1955.
7. K. A. Norton and A. C. Omberg, "The maximum range of a radar set", *Proc. Inst. Radio Engrs*, **35**, p. 4, January 1947.
8. F. H. Dickson, J. J. Egli, J. W. Herbstreit and G. S. Wickizer, "Large reductions of v.h.f. transmission loss and fading by the presence of a mountain obstacle in beyond-line-of-sight paths", *Proc. Inst. Radio Engrs*, **41**, p. 967, August 1953.
9. F. Fabbri, "Multiple paths in radio links with passive repeaters", *Alta Frequenza*, **28**, No. 3–4, p. 394, June–August 1959.
10. G. Andrieux, "Passive reflectors for radio links", *Onde Electrique*, **36**, p. 57, January 1956.
11. V. Floriani, "The Application of Reflectors in Microwave Links", Report presented at the Scientific Congress of the 3rd International Exhibition of Electronics, Atomic Energy and Radio–Television–Cinematography, Rome, 1958.
12. H. G. Booker and W. Walkinshaw, "The mode theory of tropospheric refraction and its relation to wave-guides and diffraction", "Report of a Conference on Meteorological Factors in Radio-Wave Propagation", p. 80, Physical Society and Royal Meteorological Society, London, 1946.
13. W. J. Bray and R. L. Corke, "A technique for 4000-Mc/s propagation testing for radio-relay systems", *Proc. Instn Elect. Engrs*, **99**, Pt. IIIA, p. 281, May 1952.
14. A. G. Bogle, " Some aspects of microwave fading on an optical path over sea", *Proc. Instn Elect. Engrs*, **99**, Pt. III, p. 236, September 1952.
15. R. Cabessa, "Realization of high quality radio links over sea paths in Greece", *Onde Electrique*, **35**, p. 714, August–September 1955.
16. G. Millington, "The reflection coefficient of a linearly graded layer", *Marconi Review*, **12**, No. 95, p. 140, 1949.
17. H. G. Booker, "Elements of radio meteorology—How weather and climate cause unorthodox radar vision beyond the geometrical horizon", *J. Instn Elect. Engrs*, **93**, Pt. IIIA, p. 69, March 1946.
18. M. W. Gough, " Diurnal influences in tropospheric propagation", *Marconi Review*, **21**, No. 131, p. 198, 1958.
19. E. Appleton, "The influence of tropospheric conditions on ultra-short-wave propagation", "Report of a Conference on Meteorological Factors in Radio-Wave Propagation", p. 1, Physical Society and Royal Meteorological Society, London, 1946.
20. C. S. Durst, "Radio climatology", "Report of a Conference on Meteorological Factors in Radio-Wave Propagation", p. 193, Physical Society and Royal Meteorological Society, London, 1946.
21. D. W. Heightman, "Propagation of metric waves beyond optical range", *J.Brit.I.R.E.*, **10**, p. 295, October 1950.
22. L. Essen, "A highly stable microwave oscillator and its application to the measurement of the spatial variations of refractive index in the atmosphere", *Proc. Instn Elect. Engrs*, **100**, Pt. III, p. 19, January 1953.
23. H. E. Bussey and G. Birnbaum, "Measurements of variations in atmospheric refractive index with an airborne microwave refractometer", *J. Res. Nat. Bur. Stand.* (Washington), **51**, No. 4, p. 171, October 1953.
24. O. E. de Lange, "Propagation studies at microwave frequencies by means of very short pulses", *Bell Syst. Tech. J.*, **31**, p. 91, January 1952.
25. R. L. Kaylor, "A statistical study of selective fading of super-high-frequency radio signals", *Bell Syst. Tech. J.*, **32**, p. 1187, September 1953.
26. A. B. Crawford and W. C. Jakes, "Selective fading of microwaves", *Bell Syst. Tech. J.*, **31**, p. 68, January 1952.
27. M. W. Gough, "Some features of v.h.f. tropospheric propagation", *Proc. Instn Elect. Engrs*, **102**, Pt. B, p. 43, January 1955.
28. P. Gudmandsen and B. F. Larsen, "Statistical data for microwave propagation measurements on two oversea paths in Denmark", *Acta Polytechnica*, No. 213 (Electrical Engineering Series, Vol. 7, No. 7), Copenhagen 1957.
29. Z. Jelonek, E. Fitch and J. H. H. Chalk, "Diversity reception", *Wireless Engineer*, **24**, p. 54, February 1947.
30. L. Lewin, "Diversity reception and automatic phase correction", *Proc. Instn Elect. Engrs*, **109**, Part B, 1962. (To be published.) (I.E.E. Paper No. 3584E, July 1961.)
31. R. Bateman, "Elimination of interference-type fading at microwave frequencies with spaced antennas", *Proc. Instn Radio Engrs*, **34**, p. 662, September 1946.
32. G. L. Grisdale, J. G. Morris and D. S. Palmer, "Fading of long-distance radio signals and a comparison of space- and polarization-diversity reception in the 6–18 Mc/s range", *Proc. Instn Elect. Engrs*, **104**, Pt. B, p. 39, January 1957.
33. D. G. Brennan, "Linear diversity combining techniques", *Proc. Inst. Radio Engrs*, **47**, No. 6, p. 1075, June 1957.

Manuscript received by the Institution on 20th October 1961.
(Paper No. 739.)

U.D.C. 621.386.842

The X-Ray Image Intensifier as an Inspection Tool and its Application to Stroboscopic Examination

By

C. E. PAINE†

Presented at the West Midland Section's Symposium on "New Electronic Techniques in Non-Destructive Testing" in Wolverhampton on 6th December 1961.

Summary: The x-ray image intensifier provides an image 1000 times as bright as a normal fluorescent screen with less than half the grain size. Its advantages over the fluorescent screen are discussed against a background of practical application work. The image intensifier permits several interesting techniques such as the x-ray stroboscope and the use of a closed circuit television link.

1. Description of the Image Intensifier

The principle of the x-ray image intensifier and its early application to industrial inspection problems has been discussed by Nemet and Hewett-Emmett in 1957.‡ However, it is intended here briefly to describe the intensifier and then to discuss its position in the field of industrial inspection.

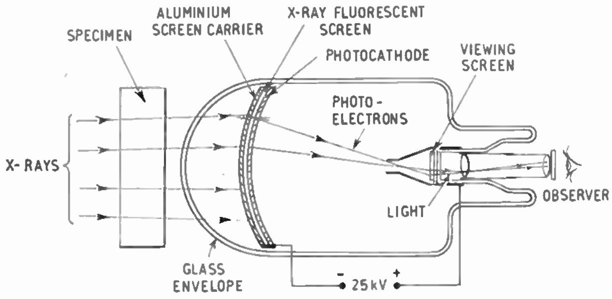


Fig. 1. Cross-section of the x-ray image intensifier tube.

The image intensifier, which is used in the same way as an x-ray fluorescent screen, consists of an evacuated glass envelope, the cross-section of which is shown in Fig. 1. The screen and photo-cathode are so shaped that the photo-electrons are focused upon the small viewing screen and the electron image reduction ratio is 9 : 1. Critical focusing of the electron beam is controlled by an electrostatic field produced by a potential applied to a conductive layer on the inside surface of the tube.

A high tension potential of approximately 22 kV is applied between the large screen (cathode) and the small screen (anode). Therefore the final image has a brightness intensification of approximately 1000 : 1 of which the image reduction produces a factor of 80. The remaining intensification between 12 and 15 is due to the acceleration of the photo-electrons along the potential gradient provided by the h.t. difference between two screens. The final image is viewed through a binocular microscope with an enlargement factor of 9 : 1.

From early experiments it was found that due to the low inherent screen blurring of the intensifier, which is approximately 0·3 mm, an enlargement technique gave considerable advantage to the definition of the final image when an ultra-fine-focus x-ray tube was used. For most industrial applications, an x-ray tube with a focal spot size of 0·4 mm is employed, as this will permit an enlargement technique of 2 : 1 to be used.

The graphs in Fig. 2 compare the sensitivity obtained with the image intensifier and a normal fluorescent screen for both aluminium and steel, when using a 150 kV constant potential unit fitted with an x-ray tube having an ultra-fine focal spot of 0·4 mm.

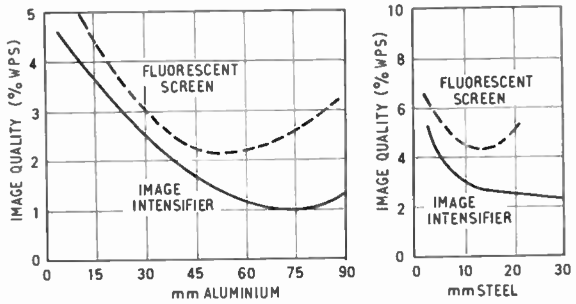


Fig. 2. Comparison of the sensitivity obtained using the image intensifier and fluorescent screens for both aluminium and steel expressed as the wire penetrameter sensitivity for different thicknesses of metal.

2. Applications of Image Intensification

During the last five years the image intensifier has been used for numerous applications and experimental work. The results of this work show clearly that whilst the image intensifier in its present form raises the level of x-ray visual examination considerably above that hitherto obtained by the standard

† Research & Control Instruments Ltd., 207 Kings Cross Road, London, W.C.1.

‡ A. Nemet and A. Hewett-Emmett. "Application of x-ray image intensification to industrial problems." *J. Brit.I.R.E.*, 17, pp. 523–7, September 1957.

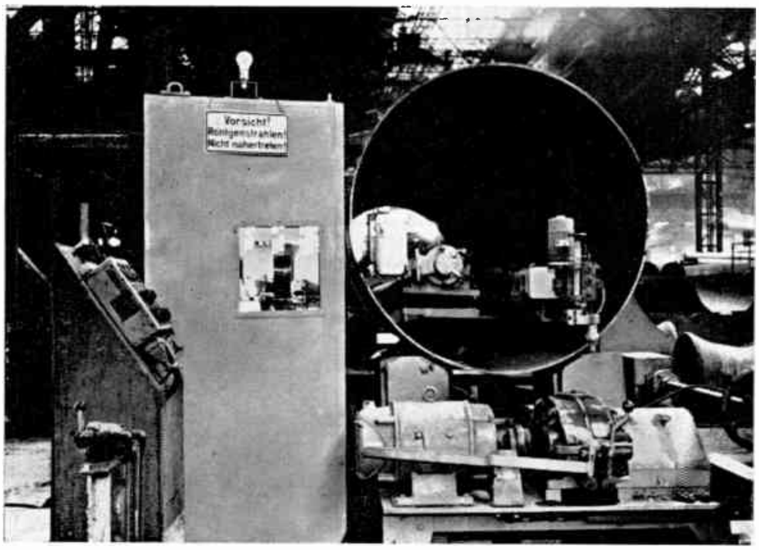


Fig. 3.
Examination of weld in large steel tube.

fluorescent screen, it does not present an image with the same amount of detail that can be obtained in a good quality radiograph. However, there is still a wide field of components that can be examined by the image intensifier such as:

- Class 2 welded fabrications
- Sealed electronic components
- High tension fuses
- Class 2 alloy castings
- Electronic valves
- Electrical components, such as switches, actuators, etc.

Before proceeding further, it should be pointed out that to obtain maximum sensitivity with the image intensifier, it is necessary that the brightness level of the whole screen should be relatively even. Should there be any extremely bright areas, these will cause halation of the screen and consequently reduce the amount of detail that can be seen in the darker areas. However, not only will the overall sensitivity be reduced, but also, damage to the photo-cathode layer can result if the intensifier is subjected to prolonged exposure to direct radiation above 100 kV. The methods employed to handle components and to avoid screen halation are described later.

From the foregoing, it will be appreciated that solid homogeneous objects of regular shape can be examined quite easily with the image intensifier as complete coverage of the screen is obtained. An example of how the intensifier may be applied to the examination of homogeneous materials is shown in Fig. 3, where the weld in a large steel tube is being examined. In this installation the x-ray tube is mounted on a rigid boom slightly longer than the welded tube. The intensifier, which is in line with the x-ray tube, is mounted in the wall of a radiation-proofed cabin which houses the operator. The control table for the x-ray unit and controls for moving the carriage supporting the steel tube are inside the cabin. Any flaws that are detected by the operator are indicated on the weld by means of a remote controlled paint spray.

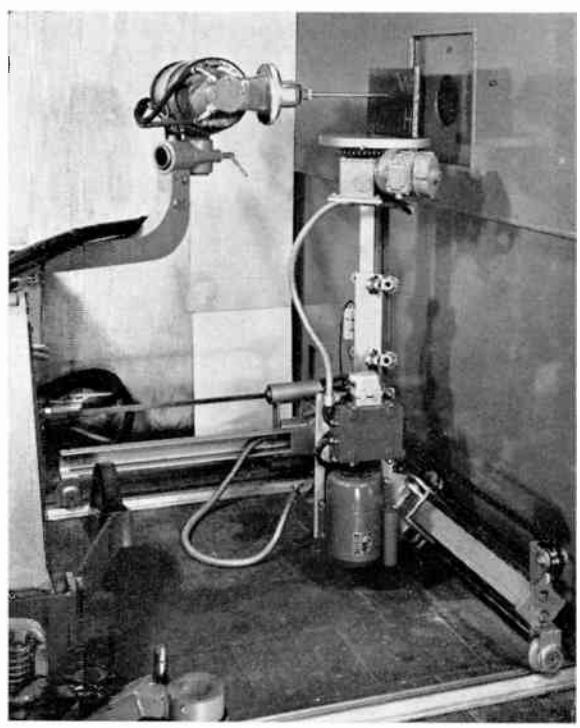


Fig. 4. Motor-driven handling gear.

When irregular shaped objects, such as castings and assemblies, which may vary in thickness and density, have to be examined, other problems arise. As mentioned earlier, it is necessary that the brightness level of the whole screen should be even. Therefore, in order to attain this the x-ray beam must be collimated so that it is approximately the size of the object under examination. To achieve this, it is necessary to have an adjustable 4-leaf diaphragm fitted to the x-ray tube. This will allow the x-ray beam to be adjusted to a suitable size for the object under examination. Radiation through holes can be eliminated by filling them with barium putty or similar material.

When an image intensifier is to be used in a laboratory for the examination of a wide range of components, it is often suitable to fit the intensifier into the wall of an existing x-ray department. Alternatively, if a separate installation is required, a cabinet constructed of lead bonded plywood may be used. In both the laboratory and the cabinet installation the image intensifier is mounted on runners which allow it a lateral movement of approximately 12 inches. Also fitted to these runners is a standard fluorescent screen and the whole installation is arranged to allow either the fluorescent screen or the intensifier to be moved into position in front of the x-ray tube.

With regard to the handling gear this should provide the components supporting table with vertical, lateral, rotation and traverse movement between intensifier and x-ray tube. These movements can be carried out by mechanical means with hand wheels mounted on the cabinet or wall just under the intensifier. However, this method is not entirely satisfactory as the operator has to use both hands for this purpose as there are usually two hand wheels on either side of him, and radiation protection of the spindle entries presents some difficulties. A more suitable method is to drive all the movements by small electric motors as shown in Fig. 4. The motors can then be controlled by means of a uni-knob switch which will allow all the movement to be controlled by one hand. An additional facility that is often desirable is the movement of the x-ray tube to allow various tube/intensifier distances to be used.

Where the routine examination of large quantities of the same component are required, an entirely different type of installation is necessary as the component has to be quickly and accurately located in front of the intensifier. The most usual way of carrying out this operation is by use of a turntable. The x-ray tube is located in the centre of the turntable facing the intensifier, which is on the outside. The components are supported in special fixtures mounted around the turntable which is rotated to bring each component in turn in front of the intensifier. An indexing mechanism is necessary for accurate alignment and a special mask shaped to the outline of the component is fitted to the intensifier.

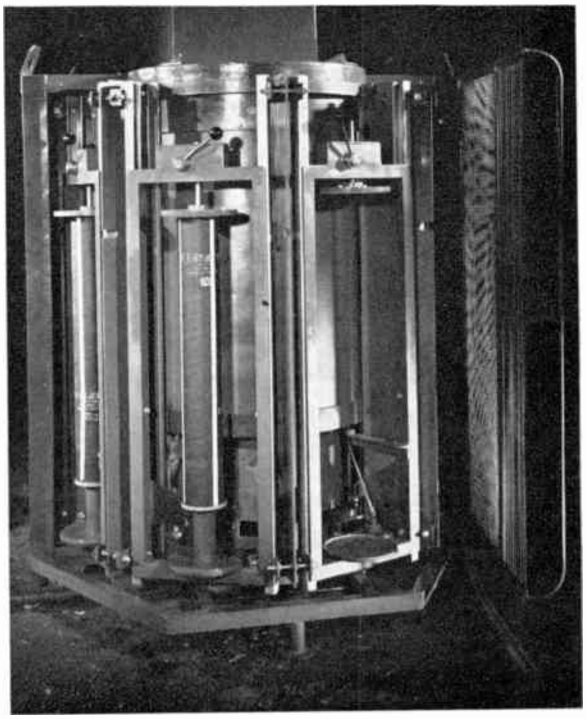


Fig. 5. Special inspection installation used for x-ray examination of cordite.

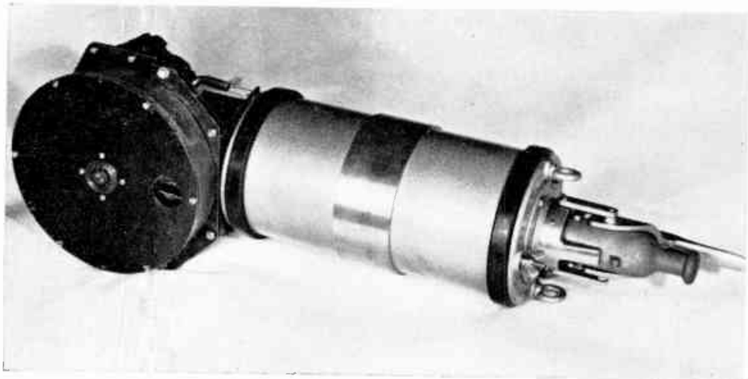


Fig. 6. Stroboscopic shutter unit fitted to x-ray tube.

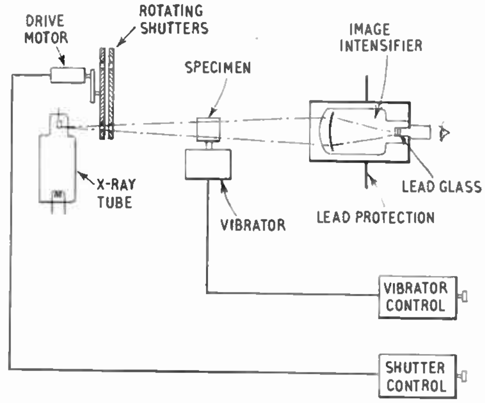


Fig. 7. Complete x-ray stroboscope.

On a correctly-designed installation, it is only a matter of minutes to change the mask and alter the fixtures for different objects. The whole installation is surrounded by a radiation-proofed shield with a loading position at the rear. Two operators are required, one to view the intensifier and one to load and unload the turntable. A typical example of such installation, which is used for the inspection of cordite, is shown in Fig. 5. To avoid direct radiation falling on the intensifier whilst the turntable is rotated, an electromagnetic shutter, which automatically covers the x-ray beam during rotation, is fitted to the x-ray tube.

In addition to the standard binocular viewer as shown on the intensifier, it is possible to fit special periscope viewers, 35 mm cameras and closed link television. The television link is particularly useful as with a suitable camera a slight increase of sensitivity can be obtained and the monitor screen can be viewed by several persons at the same time.

3. X-ray Stroboscopy

During the last few years the rapid increase of environmental testing has brought with it a number of problems for the inspection engineer. One of these problems was to find out what is happening inside a sealed component whilst undergoing an environmental vibration test. To cut away part of the case or replace with a transparent window would completely alter the resonant characteristics of the component. Therefore, it was necessary to devise a means for inspecting the inside of the component without altering the construction. As the image intensifier was already being used for the static inspection of such components, experiments were carried out to see if it could be adapted for stroboscopic application. From these tests it was found that the intensifier could present a satisfactory image if the x-ray beam was pulsed at a frequency within ± 4 c/s of the vibration frequency. It was therefore decided to adapt a standard 150 kV constant potential x-ray unit to supply a pulsed x-ray source from 0–150 c/s. This could have been achieved by an electronic h.t. switch unit or a mechanical rotating shutter. The latter method was employed as it was considered to be less costly and in any case, there were no 150 kV triode control valves available.

The final construction of shutter consisted of two contra-rotating discs, each having four apertures. These discs are driven by a split-field d.c. servo motor, which retains a constant speed to within ± 1%. The finished shutter unit is shown fitted to a standard 150 kV x-ray tube in Fig. 6.

The results obtained from this apparatus have proved to be most satisfactory. Although the initial installation had separate controls for the vibrator and shutter unit, as shown in the schematic diagram in Fig. 7, it was quite easy to control and carry out a resonant search. The effective range of the shutter is between 35–150 c/s. However, it can be used on multiples of this frequency up to 600 c/s. Above this frequency the amplitude of components is too small to be visible on the intensifier. A further development of the apparatus will be the incorporation of a feedback system from the vibrator control to keep the shutter speed in synchronism with the vibrator. This would simplify the control of the apparatus.

Although the application of this apparatus is still in development stage, the work carried out so far has proved to be of immense value for investigation into environmental testing problems. A further development contemplated is the use of the stroboscopic shutter and image intensifier for the investigation of the vibration in complete structures, such as aircraft, missiles, etc. For this use special handling gear would be required and a closed circuit television camera chain will be fitted to the intensifier. Although such an installation would be very costly, it is felt that the information gained and time saved would be worthwhile.

Manuscript received by the Institution on 7th March 1962 (Paper No. 741).

U.D.C. 534.23: 620.179

Detecting Flaws in Steel Tube or Bar with a Rotating Coil in a Magnetic Saturating Field

By

W. H. BAKER (*Associate*)†

Presented at the West Midland Section's Symposium on "New Electronic Techniques in Non-Destructive Testing" in Wolverhampton on 6th December 1961.

Summary: The equipment is unusual in that either eddy current or magnetic flux leakage is employed for flaw detection. A planetary scanning coil is rotated at high speed around the periphery of the tube, the incremental permeability of which is reduced to approaching unity by using an intense magnetic field. Flaws 0·002 in. deep are detected under suitable conditions. The relative results of the two techniques are discussed against a background of laboratory trials using this equipment.

1. Introduction

Present methods of non-destructive testing of steel tube by eddy current techniques usually use an encircling pick-up coil through which the tube is threaded. An alternative, known to give greater sensitivity in the case of non-ferrous metals, is to use a small pancake coil in lieu of the encircling coil.‡ Complete coverage of the tube circumference is usually obtained by spinning the tube under the pancake coil, whilst the latter is traversed along the length of the tube, so obtaining a spiral scan.

The work to be described was undertaken to assess the sensitivity and general suitability of the scanning coil technique for detecting predominantly longitudinal flaws in long lengths of steel tube and bar. An essential feature of the investigation was that there should be not less than 0·080 in. clearance around the tube surface.

It was decided that a convenient pick-up coil size was $\frac{1}{4}$ in. by $\frac{1}{8}$ in. as this should resolve a fine longitudinal hair crack and yet have adequate mechanical strength. Because the tubes to be tested were too long to be rotated satisfactorily, the pick-up coil had to be put into orbit around the tube. Mechanical considerations determined that the speed of rotation should not be more than 2000 rev/min.

It was considered necessary to reduce the incremental permeability of the steel under test to a value approaching unity by a state of magnetic saturation. Experimental work carried out showed that a field strength of approximately 2000 oersteds was needed for this purpose (see Fig. 1).

2. Eddy Current Flaw Detection

An e.m.f. can only be induced into the pick-up coil if there is a suitable source of electromagnetic energy. A well-known technique is to have a high frequency source in the form of an exciting coil which is coupled to the pick-up coil via eddy currents induced into the workpiece (i.e. the tube). Discontinuities such as cracks and voids, or inhomogeneities such as inclusions, divert the normal eddy current path so altering the coupling between exciting coil and pick-up coil. This brings about a change of e.m.f. in the pick-up coil and this is used to indicate the presence of a defect.

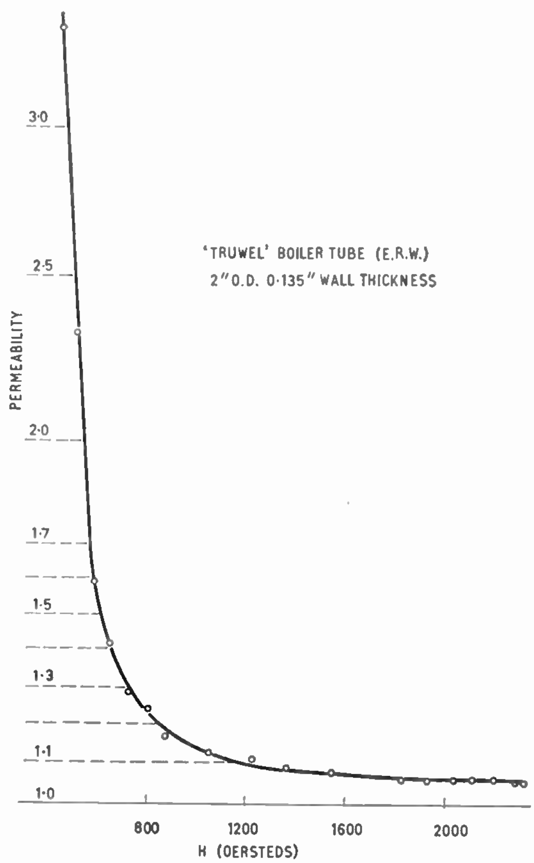


Fig. 1. Measurement of incremental permeability for changes in field strength.

† Tube Investments Technological Centre, Walsall, Staffs.
‡ "New developments in non-destructive testing of precision steel tubes", *Instrument Practice*, 15, No. 5, pp. 553–6, May 1961.

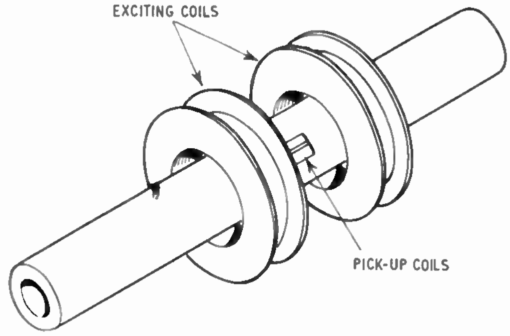


Fig. 2. Double exciting coil arrangement for inducing eddy currents.

For example eddy currents induced circumferentially around the tube will be diverted to maximum effect by longitudinal type defects. This diversion will produce a radial component of flux which is readily picked up by a rotating pancake coil. Eddy currents can be circumferentially energized by means of a simple exciting coil encircling the tube.

An experiment was mounted to investigate the sensitivity and explore the geometrical parameters of an eddy current system employing a single encircling exciting coil and a rotating pancake pick-up coil. For convenience a brass tube was used instead of steel as this avoided the complication of having a polarizing coil to saturate the steel. Thus the relative physical disposition of the coils was established and it was demonstrated that cracks 0·004 in. deep could be detected in brass.

An alternative way of inducing circumferential eddy currents into the tube is to use two opposingly wound exciting coils spaced side by side with the pick-up coil between them (see Fig. 2). Thus the ampere-turns of the exciting system is doubled and the flux gradient in the vicinity of the pick-up coil is increased. Although no laboratory experiments were tried using the double exciting coil, it was decided to include this feature in the finished model because of its intrinsic advantage for long longitudinal flaws as against minor (short) blemishes.

The pick-up coil itself was also designed as a double unit. By connecting the coils in opposition a null balance can be obtained under conditions of no flaw. The field disturbance produced by a defect is detected by each coil successively, so producing an out-of-balance signal that is a differential of the radial component of the flux pattern produced by the presence of the defect. The e.m.f. generated in the pick-up coils constitutes the flaw signal and it appears as a modulation on a carrier whose frequency is that of the eddy current generator.

3. The Eddy Current System Used

A schematic representation of the eddy current system used is shown in Fig. 3.

The steel tube is heavily saturated by means of the polarizing solenoid indicated. This solenoid is supplied with d.c. from a power supply. Within the saturating field there are the two exciting coils and the transducer comprising the balanced pair of pick-up coils.

The exciting coils are supplied from a variable frequency oscillator capable of supplying 150 milliamperes into an exciting coil load of 3 kilohms at frequencies up to 100 kc/s. The oscillator has two other sinusoidal outputs: one which is used as a reference signal, and the other which can be varied in phase with respect to the reference signal.

The outputs from the rotating pick-up coils are fed through slip rings (not shown) to a balancing amplifier which provides the null balance conditions in the absence of flaws. The out-of-balance signal is then fed through a high-gain standard laboratory amplifier (up to 94 dB) and thence to the phase detector.

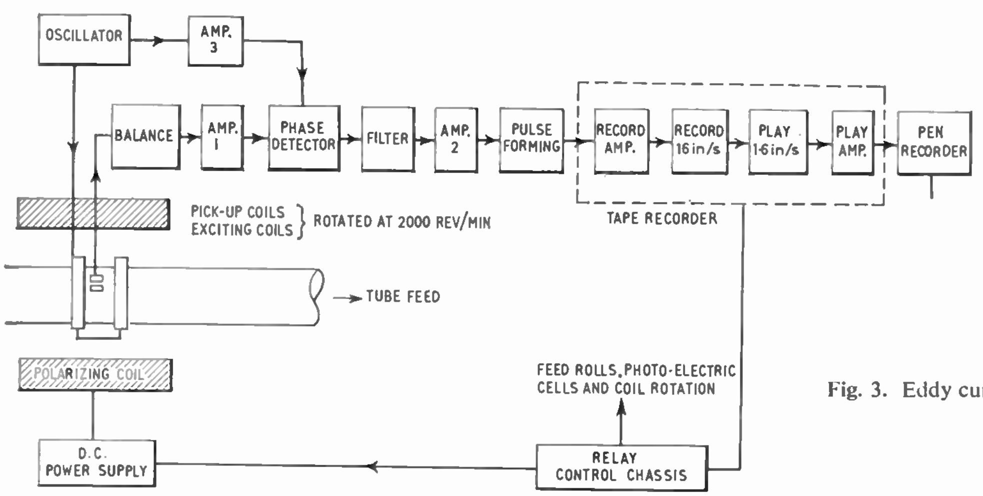


Fig. 3. Eddy current diagrammatic layout.

The out-of-balance flaw signal is a modulated waveform at oscillator frequency and is compounded of phase and amplitude modulation. Consequently the detector is designed to provide a means of phase discrimination as well as acting as a detector: hence the reference and variable phase voltages fed in from the oscillator via the power amplifier.

Discounting inherent noise in the system, the demodulated signal consists of a combination of wanted signals caused by flaws and unwanted signals caused by non-uniformity of tube material under the pick-up coil as it orbits around the tube. A frequency discriminating filter having band-pass characteristics is tuned to the characteristic frequency of the flaw which in the first instance was found empirically.

In practice the flaw signal has a frequency range of 300–900 c/s dependent mainly on the diameter of the tube inspected. As no pen recorder capable of coping with this frequency was to hand, the flaw signals were amplified and pulse formed so as to be suitable for recording on magnetic tape. By playing back at a speed reduced by 10 : 1, the tape recorder output can be fed into a pen recorder having a maximum frequency response of 90 c/s.

Fig. 4. Flaw signals before and after pulse forming.

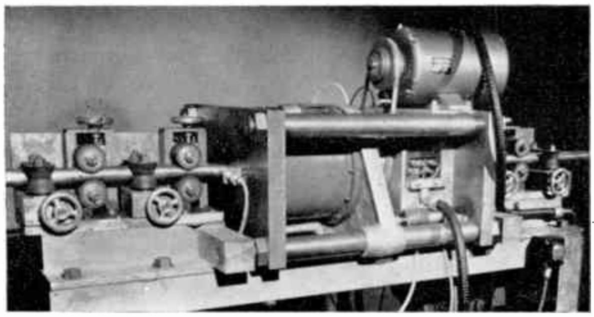


Fig. 5. Inspection head complete with feed rolls.

Alternatively the output from the pulse-forming unit can actuate a standard flaw alarm level monitor of conventional design. The lower trace of Fig. 4 shows the resultant signal as reproduced by the pen recorder after the signal shown in the top trace has been passed through the pulse-forming unit.

4. Description of Apparatus

The electronic equipment (Fig. 3) is housed in a single cubicle and consists of a number of rack mounted panels. This permits certain standard equipment such as amplifiers and a tuneable bandpass filter

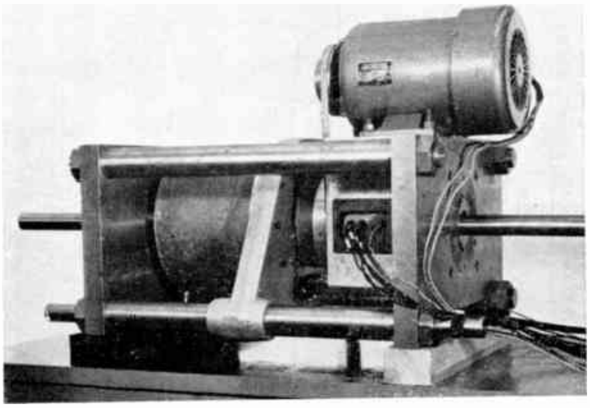


Fig. 6. Inspection head.

to be incorporated with other specially designed units such as the oscillator and phase detector.

The tape recorder/pen recorder with associated amplifiers is mounted on a self-contained removable trolley which can be unplugged from the cubicle so that the tape can be transcribed at leisure, away from the main plant.

The pick-up, exciting and polarizing coils are all built into an inspection head which is located between two sets of feed rolls (see Fig. 5). The inspection head alone is shown in Fig. 6. It will be seen that the tube is threaded through the inspection head and that an electric motor, shown at the top of the main assembly, drives the central rotating unit by means of a belt. This motor was found to be in the leakage field of the magnetic polarizing system and has since been removed and placed underneath the main assembly.

The construction of the main assembly is illustrated in section (Fig. 7). A tube is shown in position threading through the axial orifice. The assembly can be considered as two portions: stator and rotor. The stator consists essentially of an air-gapped yoke which carries the polarizing coil and the bearings displaced about the air-gap. The rotating pick-up coil and exciting coil are in this 2-inch air-gap. The flux path for the magnetic saturating of the tube is from the polarizing coil, through the air-gap (and of course the steel tube if in position), through a central pole piece adjacent to the tube, around the steel end-plate back through the four steel supporting arms, returning to the polarizing coil via the other end-plate and central pole piece between tube and polarizing coil. The rotor runs on the ball bearings shown, and is in itself in two parts. The permanent half contains the slip rings which connect the rotating coils to the stator housing via brush gear, and the crowned pulley which takes the belt drive from the drive motor (see Figs. 5 and 6). The detachable half sits in the 2 in. air-gap between the central pole pieces. It houses the pick-up coils and the encircling eddy current exciting coils (the latter do

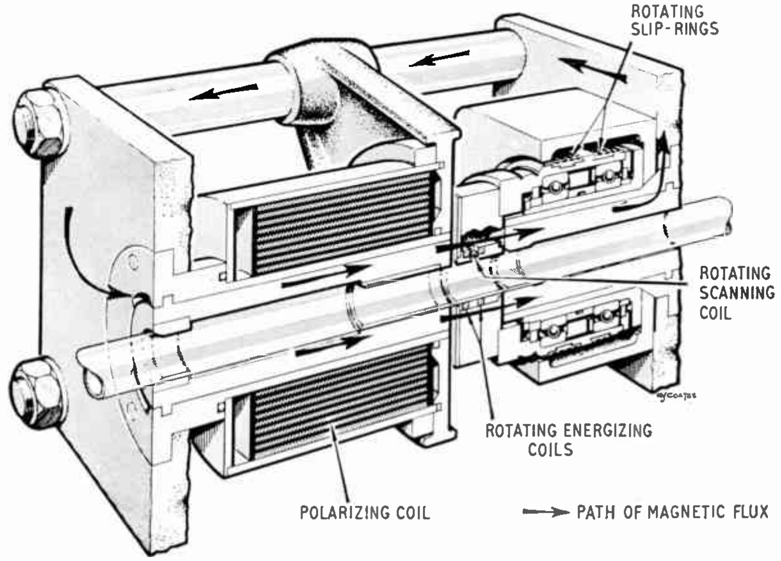


Fig. 7. Section view of inspection head.

not of course need to be rotated; this is a functional convenience) and is so constructed that it connects to the permanent half of the rotor by plug and socket (see Fig. 8). It is bolted into position. This arrangement facilitates changing the coil units with varying tube diameter. There is access to the coil unit because the whole polarizing coil assembly can be moved up 2 in. by two air cylinders (not shown), so taking up the gap between its end-cheek and the end-plate of the main assembly. When changing tube size it is also necessary to change the two detachable pole pieces which lie axially between the air-gap and the end-cheeks of the main assembly.

The field strength in the region of the pick-up coil is approximately 2000 oersteds. This is derived from the polarizing coil which is spaced layer wound with 3000 turns of 16 s.w.g. wire and cooled by forced oil. The oil is in turn water cooled by a simple heat exchanger system which limits the maximum temperature rise to 18° C above ambient when the coil is supplied with 7 amperes d.c.

5. Results Obtained Using Eddy Currents

To assess the sensitivity of the equipment, tests were made using milled artificial defects in a steel tube honed internally and externally. Equipment performance was studied for both transverse and longitudinal defects. The entire experiment was performed using a single eddy current exciting coil system and two pick-up coils in opposition (as described in section 3). The experiment was then repeated with a two exciting coil system, no attempt was made to balance the coils exactly.

Four flaw depths were used: 0·002 in., 0·005 in., 0·010 in., and 0·015 in.; all approximately 0·006 in. wide. The rotating coil was approximately 0·080 in. from the outer surface of the tube. Figure 9 shows typical traces observed on an oscilloscope connected to the output of the filter (see Fig. 3). From this it will be seen that an 0·005 in. longitudinal flaw is just about repeatedly discernible through the general noise on the trace. This figure of 0·005 in. is quoted as the maximum sensitivity of the system. It is intended as a reference for comparison purposes within the framework of this experiment and that described in Section 9 rather than an absolute measurement. On this basis the Table 1 has been derived from the records taken.

For longitudinal flaws both coil configurations gave comparable sensitivities, although it had been expected that the double coil would be a little better because of the increased flux gradient. For transverse flaws the double coil failed to produce a flaw signal even at 0·015 in. flaw depth.

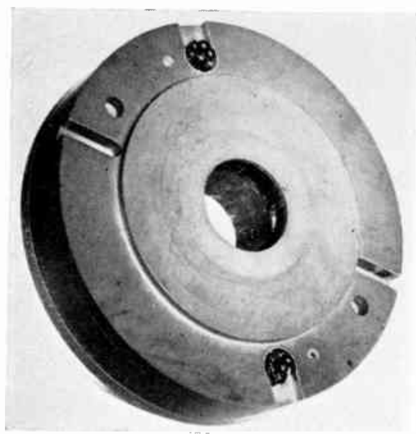


Fig. 8. Rotating coil assembly housing, exciting and pick-up coils.

Table 1

Sensitivity tests using eddy currents (50 kc/s) on mild steel tube.

Exciting Coil Configuration	Flaw Type	
	Longitudinal	Transverse
Single	0·005 in.	0·005 in.
Double	0·005 in.	None detected

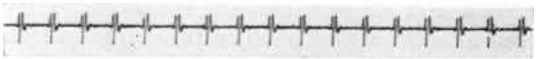
Air-gap 0·080 in. Saturation field strength approximately 2000 oersteds.

As expected the circumferential method of inducing eddy current detects longitudinal flaws better than transverse ones. Although Table 1 shows that both types of flaw are discernible when 0·005 in. deep, the records clearly indicate that the signal-to-noise ratio improves markedly with flaw depth when the flaw is longitudinal, but remains substantially constant from 0·005 in. to 0·015 in. when the flaw is transverse. It appears that no indication of flaw depth is therefore possible when using circumferential eddy currents to detect a transverse flaw with a scanning pancake coil.

(*a*) 0·002 in. longitudinal.

(*b*) 0·005 in. longitudinal.

(*c*) 0·010 in. longitudinal.



(*d*) 0·015 in. longitudinal.

Fig. 9. Flaw signals. Eddy currents. 1 exciting coil. Longitudinal flaws.

6. Magnetic Flux Leakage Flaw Detection

A rotating coil in a saturating field can be used as a flaw detector because of the induced changes in e.m.f. every time the coil passes through the local field around the defect. The principle of magnetic flux leakage crack detection can be seen from Fig. 10. In Fig. 10(*a*) is shown a section of magnetic lines of flux through a steel tube suitably magnetized but not magnetically saturated. There will be zero induction into the pick-up coil if the tube wall is sound because the lines of flux are parallel to the pick-up winding. Figure 10(*b*) shows what happens if there is a mid-wall void. A small disturbance in the flux pattern occurs which will be picked up by coil *P*. This much has been known and used for many years: it is, however, a relatively insensitive system because the steel itself acts as a magnetic shield between the void and the pick-up coil. However, this machine has been developed to produce a state of magnetic saturation in the steel tube: this is shown in (*c*). If now there is a mid-wall void, the effect of the flux disturbance on the coil *P* is greater, as shown in (*d*). This is because the wall being saturated now is effectively non-magnetic to the small variations due to the flaw and no longer shields the void from the pick-up coil.

7. The Magnetic Flux Leakage System Used

The flux disturbance produced by the flaw induces an e.m.f. into the rotating coil (which may be single or twin) each time the coil passes through it. This e.m.f. is in the form of simple pulse having a main frequency component, in this case, of between 300 and 900 c/s. This pulse is of greater amplitude than in the case of eddy currents and so only one amplifier is required before filtering and pulse forming prior to recording. (See Fig. 11.) The recorder and filter are used as in the same way as described for eddy currents (Section 4).

It will be noted that the flux leakage system does not require a radio frequency carrier. This means that the oscillator and the demodulation circuits required for eddy currents are unnecessary as of course are the energizing coils in the rotary head.

8. Results Obtained using Flux Leakage

The experiments described in Section 6 were repeated using the equipment as a flux leakage flaw detector.

The same steel tubes having the same longitudinal and transverse flaws were employed, and as before, four sets of results were obtained. Energizing coils were not of course used, but the rotating transducer was connected firstly as a single pick-up coil, and secondly as a double coil, the windings being connected in series opposition.

The basis for judging sensitivity is as in the eddy current experiment (Section 5) and it is thought a direct comparison can therefore be made. Figure 12 shows the output wave from filter for comparison with Fig. 9.

The use of a double pick-up coil was an attempt to reduce the background noise by connecting the coils in opposition in the hope that their background e.m.f.s due to standing flux patterns would be bucked out, whilst the wanted e.m.f.s due to flaws would be unimpaired because the coils intercept the flaw's

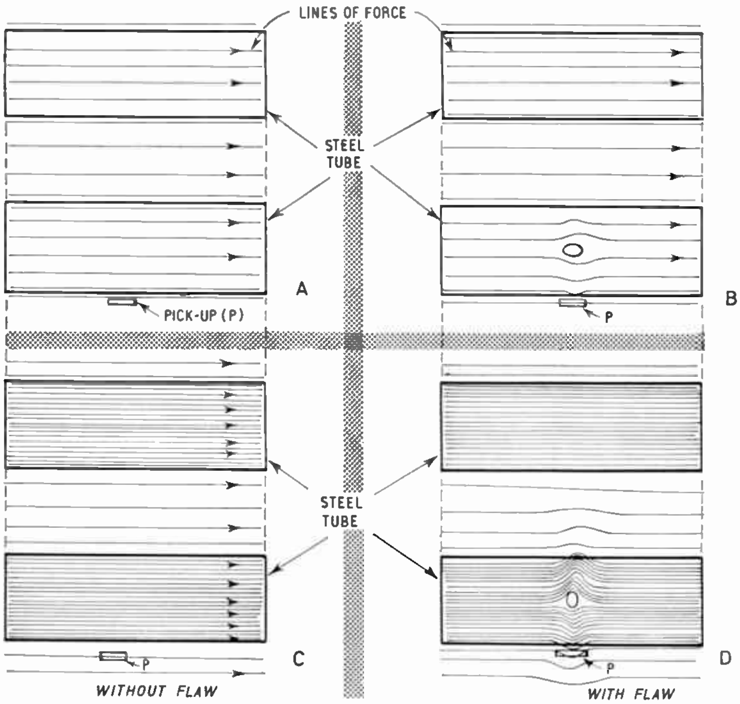


Fig. 10. Principles of magnetic flux leakage flaw detection.

Table 2

Sensitivity tests using magnetic flux leakage on same mild steel tube as referred to in Table 1

Pick-up Coil Configuration	Flaw Type	
	Longitudinal	Transverse
Single	0·002 in.	0·002 in. (probably better)
Double	0·005 in.	0·002 in.

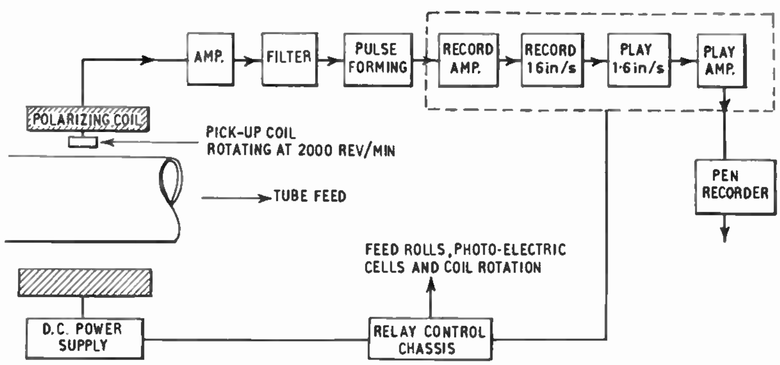
Air-gap and polarizing current as for Table 1.

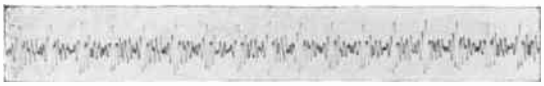
disturbance field in succession. In fact Table 2 shows the single coil to be superior. This is because the two coils were located one behind the other in the plane of rotation, and so picked up out-of-phase signals from general flux inequalities not caused by flaws, which increased rather than diminished the background noise.

As expected the longitudinal polarization of the magnetic flux in and around the tube makes the device more sensitive to transverse flaws than to longitudinal ones. In fact the record for a 0·002 in. transverse flaw (Fig. 13) is such that it is reasonable to suppose that a 0·001 in. flaw could also be detected at 0·080 in. air-gap.

During experiments it was noticed that the distance between the tube surface and the coil had a pronounced effect on the amplitude of the signal derived from any particular flaw. This effect was studied using a steel tube having two artificial longitudinal defects 0·010 in. and 0·020 in. deep respectively. The results are shown

Fig. 11. Magnetic flux leakage diagrammatic layout.





(*a*) 0·002 in. longitudinal.

(*b*) 0·005 in. longitudinal.

(*c*) 0·010 in. longitudinal.

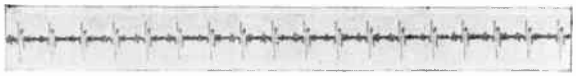
(*d*) 0·015 in. longitudinal.

Fig. 12. Flaw signals. Magnetic flux leakage. Longitudinal flaws.

(*a*) 0·002 in. transverse.

(*b*) 0·005 in. transverse.

(*c*) 0·010 in. transverse.



(*d*) 0·015 in. transverse.

Fig. 13. Flaw signals. Magnetic flux leakage. Transverse flaws.

graphically in Fig. 14. Two points of interest emerge.

(1) The signal-to-noise ratio: at 0·1 in. clearance a flaw 0·010 in. deep is still reasonably clear of the noise level (about 11 dB).

(2) The susceptibility of flaw depth measurement to variations in distance between tube surface and pick-up coil: a relative movement of about 0·020 in. between coil and tube is sufficient for a 0·020 in. deep defect to appear to be 0·010 in. deep (or vice versa).

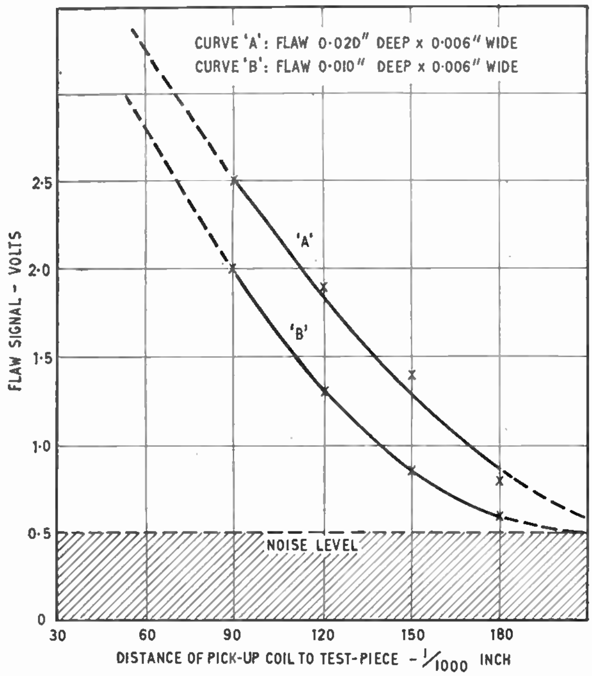


Fig. 14. Variations in signal strength with variations in distance between tube surface and pick-up coil.

9. Conclusions

Under closely controlled conditions the sensitivity of a rotating coil $\frac{1}{4}$ in. by $\frac{1}{8}$ in. having an 0·080 in. air-gap between pick-up and workpiece is such that it will detect a longitudinal flaw 0·005 in. deep by $\frac{1}{2}$ in. long by 0·006 in. wide in mild steel tube, when operated as an eddy current flaw detector.

Under similar conditions, but operated as a flux leakage flaw detector, a flaw 0·002 in. deep by $\frac{1}{2}$ in. long by 0·006 in. wide can be detected. Flux leakage flaw detection is easier to apply and requires less apparatus than the corresponding eddy current system.

Experimental work on flux leakage flaw detection shows that the signal received from a given depth of flaw varies appreciably with changes in tube surface/pick-up coil distance. If flaw depth is to be assessed consistently under practical working conditions of mechanical tolerances, a correction for tube surface/pick-up coil variations must be made. There are well-known electronic techniques which would permit this to be carried out automatically.

Manuscript first received by the Institution on 6th December 1961 and in final form on 21st May 1962. (Paper No. 742.)

DISCUSSION ON

"New Electronic Techniques in Non-Destructive Testing"

In the Chair: Mr. P. Huggins (*Member*)

Mr. B. L. Davies: I would like to refer to the part of the paper by Messrs. Kay, Whipp and Bishop† dealing with the improvement of signal to noise ratio. When inspecting material continuously in line, defect echoes often persist at roughly the same range due to the defects being of an extended nature in the line direction. Although the defect signal is different for each pulse since it is actually from a different point of the material can the fact that it appears at the same range be used to improve the signal to noise ratio?

Messrs. L. Kay, E. Whipp and M. J. Bishop (*in reply*): In any echo-location system the echo background rate can be improved only if the background changes more often than the echo from a particular discontinuity in the medium. Movement of the ultrasonic transducer over the surface of a metal produces a change in the background signal since this is produced by a relatively large number of small scatterers. If at the same time, the echo from a defect changes less than the background, an improvement in the echo/background ratio is possible. The example described by Mr. Davies may exhibit these features since the range remains constant. The fact that the range of the defect is constant cannot, by itself, be used to improve the echo/background ratio.

Mr. J. K. Gessler: May I compliment Mr. Chattaway‡ on the elegance of his method of overcoming the practical inconveniences of orthodox immersion techniques.

Was it necessary to have a four pen recorder? It would appear that neglecting the distance marked and by omitting back each amplitude (which seemed of little advantage), a single pen should in many cases give all the information required.

Although Fig. 3 is diagrammatic, the lower portion of Fig. 6 shows the record obtainable in practice from one pen. It gives the length of each defect and its distance from the edge of the bar. Since the electronic circuit makes it possible to determine the level below which defects are not to be recorded, there seems little purpose in having also a shaded picture of the fault echo amplitude.

A similar ultrasonic head mounted at right angles to the first, both working in short bursts out of phase, could feed two pens and hence give the co-ordinates of three dimensional position of the defect.

Mr. M. D. Chattaway (*in reply*): In order to locate defects in the bar it is necessary to record the distance along the bar. Back echo amplitude is also an important factor in assessing the degree of contact (bad surface conditions etc.) and to find areas of high attenuation, porosity etc.

The purpose of the shading on the recording is twofold, firstly it makes defects recorded on the flow-channel more obvious, and secondly it provides a constant zero reference that makes d.c. drift in the recorder unimportant.

As Mr. Gessler remarks, two heads can be mounted at 90 deg to each other to give the radial range from the transducers, and enable the position of the defect to be "fixed".

Mr. J. Jones: In our experience "Polycell" acted as a better couplant than either oil or grease, and little difficulty was experienced in maintaining contact over long lengths (10 ft) with hand-held probe. Grinding of bars prior to testing obviously governed the efficiency.

The self-contained "unit" described by Mr. Chattaway seemed to work very well on the 5 in. diameter bar, but how practical is the unit when a quantity of bars have to be tested (e.g. 20 or 30 of 2–3 in. square); does the manipulation from bar-to-bar necessitate "setting-up" each time?

Is it as efficient, when suitably adapted, on a square bar?

Mr. Chattaway (*in reply*): It is generally acknowledged that the immersion system gives a more consistent coupling ultrasonically. The use of "Polycell" results in rapid corrosion of mild steel.

A great number of bars have been tested using the test head and setting up is necessary only once or twice per day, using a test block to check sensitivity etc.

The unit could be adapted to be just as efficient on square bar, except that the corners of the bar would not be tested without the use of special techniques.

Mr. B. P. Hayward: Do any of the authors envisage the possibility of the use of any ultrasonic technique to determine whether or not an actual weld "nugget" has been produced between two sheet components, the thickness of which may be between the limits 0·020 in. and 0·1 in.? Such an inspection device should be capable of differentiating between two separate sheet components fused into intimate contact, and the creation of an actual fused square joining the two sheets together.

Mr. Chattaway (*in reply*): An ultrasonic method could be developed to locate areas of non-fusion, possibly using a transmitting probe on one side of the sheet, and a receiver probe on the other. Severe attenuation of the transmitted signal occurs when a non-fused area is traversed. For the thicknesses quoted it will be necessary to use high frequency probes in order that the thickness shall have a dimension equal to several wavelengths of the transmitted energy.

Messrs. Kay, Whipp and Bishop (*in reply*): In reply to Mr. Hayward's question, the authors are not sufficiently familiar with the process of welding thin sheets to give a definite answer. It would seem that a through transmission technique would be more suitable than a reflection

†L. Kay, E. Whipp and M. J. Bishop, "The physical factors affecting the reliability of ultrasonic non-destructive testing", *J. Brit.I.R.E.*, **23**, pp. 373–80, May 1962.

‡M. D. Chattaway, "Automatic charting of ultrasonically detected flaws in bar", *J. Brit.I.R.E.*, **23**, pp. 281–5, April 1962.

method, but if the acoustic impedance and absorption of the two forms of fusion are different some information regarding the weld may be possible.

Mr. B. P. Hayward: In the absence of other suitable inspection techniques, x-rays are often used for the evaluation of the quality of resistance welds between sheet components. However, considering single spot welds, the image obtained may only be approximately ⅜ in. diameter and it is defects in this area which are being sought. Such radiographs can consequently only be accurately interpreted by highly skilled operators with many years experience. Where the volume of work is such that numerous operators are required to deal with it, some system of improvement in the definition of image is required together with a standardization of interpretation so that less skilled people are permitted to make decisions. Could Mr. Paine† make any constructive comments regarding such problems?

Mr. C. E. Paine (*in reply*): The problems of non-destructive testing the resistance welds between sheet components is a difficult one. Experience has shown that the type of defects that occur in resistance welds are very minute cracks and, in the case of light alloys, copper segregation. Therefore, to-date the examination of this type of weld has of necessity been with x-rays using a fine grain film. As previously pointed out, due to the fact that the intensifier gives a sensitivity only approaching that of radiographs on medium grain film, it is not suitable for use in the examination of resistance welds.

†C. E. Paine, "X-ray image intensifier as an inspection tool and its application to stroboscopic elements", *J. Brit.I.R.E.*, 24, pp. 73–6, July 1962.

Mr. S. Goldberg: Can Mr. Paine give the "figure of merit" in terms of minimum detectable change of density for the optical image intensifier he has described, and can he compare this with the latest commercial electro-luminescent image intensifier?

Mr. Paine (*in reply*): The x-ray image intensifier gives a "figure of merit" of between 3–4% for the minimum detectable change of density as compared with the infrared image amplifier which is below 0·5%.

Dr. D. L. Emberson: Is the limit of sensitivity with the intensifier system Mr. Paine has described set by the photon noise of the incident x-ray photons reached? If not, is there sufficient light produced by a single x-ray photon incident on a fluorescent screen to allow the use of efficient coupling optics between the fluorescent screen and the photocathode of an image intensifier so that one x-ray photon produced at least one photoelectron from the photocathode? If this is possible, the use of one of the presently available multi-stage image intensifiers, for example of the transmission secondary emission type, capable of detecting single photoelectrons might improve the sensitivity of the system.

Mr. Paine (*in reply*): The limit of sensitivity of the intensifier is set by the photon noise of the incident x-ray photons. Due to the fact that less than 1% of the x-ray quanta are absorbed by the fluorescent screen, under certain conditions, there will not be sufficient light produced by a single x-ray quanta so that photoelectrons are produced. Therefore, in this case the use of multi-stage image amplifiers in conjunction with the x-ray image intensifier would not improve the sensitivity of the system.

Proceedings at the Symposium on
"New Electronic Techniques in Non-Destructive Testing"

The following papers read at the above meeting held in Wolverhampton in December 1961 have now been published:—

"The Physical Factors Affecting the Reliability of Ultrasonic Non-Destructive Testing—A Review of Current Research"—L. Kay, B.Sc., E. Whipp, B.Sc., and M. J. Bishop, B.Sc. (*May*)

"Automatic Charting of Ultrasonically Detected Flaws in Bar," M. D. Chattaway (*April*).

"Detecting Flaws in Steel Tube or Bar with a Rotating Coil in a Magnetic Saturating Field"—W. H. Baker (Associate) (*July*).

"X-ray Image Intensifier as an Inspection Tool and its Application to Stroboscopic Examination"—C. E. Paine (*July*).

"Automatic Evaluation of Defect Severity by Shape and Size"—D. R. Aldridge-Cox (*April*).

Reprints are available from the Institution price 3s. 6d. each or 15s. for the set of five papers plus discussion.

APPLICANTS FOR ELECTION AND TRANSFER

The Membership Committee at its meeting on 26th June last recommended to the Council the election and transfer of 18 candidates to Corporate Membership of the Institution and the election and transfer of 27 candidates to Graduateship and Associateship. In accordance with Bye-Law 21, as adopted at the Special General Meeting held on 23rd May 1962, the Council has directed that the names of the following candidates shall be published under the grade of membership to which election or transfer is proposed by the Council. Any communications from Corporate Members concerning these proposed elections must be addressed by letter to the Secretary within twenty-eight days after the publication of these details.

CORPORATE MEMBERS

Transfer from Associate Member to Member

MURPHY, Timothy Joseph, C.G.I.A. *Mount Merrion, Co. Dublin.*

Direct Election to Associate Member

BANFIELD, Bernard John. *Leighton Buzzard, Bedfordshire.*
BARKER, Cyril Glyn. *South Ruislip, Middlesex.*
BINDON, Douglas George. *Nottingham.*
DURHAM, Eric, B.Sc. *Wirral, Cheshire.*
GOODACRE, Jack, B.Sc.(Eng). *Baughurst, Hampshire.*
*McMULLAN, Walter Joseph. *Bangor, Northern Ireland.*
SYLVESTER-BRADLEY, Major, J. T. R., M.A., R.Sigs. *Catterick Camp.*

Transfer from Graduate to Associate Member

BARRETT, Brendan. *Dublin, Eire.*
BELL, Duncan. *Dunbar, East Lothian.*
BOAKES, Bernard Frederick. *South Hornchurch, Essex.*
CHALLENGER, Michael Trevor. *Cheltenham, Gloucestershire.*
HURST, Sydney. *Salisbury, South Australia.*
LANGTON, Victor Rae Musk. *Bexley, Kent.*
POOLE, Lloyd William. *Little Chalfont, Buckinghamshire.*
YOUNG, Lester Harold. *Freetown, Sierra Leone.*

Transfer from Student to Associate Member

BROWN, Norman George. *St. Albans, Hertfordshire.*
KENNETT, Barrington George. *Paignton, Devon.*

NON-CORPORATE MEMBERS

Direct Election to Associate

BROWN, William Thomas Royston. *Weybridge, Surrey.*
GODFREY, Raymond William. *Gillingham, Kent.*
HARRISON, Leonard William. *Wainuiomata, New Zealand.*
JORDAN, Edward James. *Marlow, Buckinghamshire.*
LESTER, Cecil Lloyd. *Croxley Green, Hertfordshire.*
MACKENZIE, George Herbert. *Causeway, Rhodesia and Nyasaland.*
MOBSBY, Raymond Bright. *Brookwood, Surrey.*
THORNTHWAITE, Eric James. *Thurso, Caithness.*
GILL, Brian Sidney. *Winchester, Hampshire.*
KENNARD, David Edward. *Shepperton, Middlesex.*
LEAMAN, Kenneth Edwin. *Malvern Link, Worcestershire.*
MARTIN, Earl Hubert. *Ottawa, Ontario, Canada.*
MUNN, Major David Charles, M.A.(Cantab). *Bushey, Hertfordshire.*
PATTERSON, Donald, B.Sc. *Southminster, Essex.*
PLENTY, Adrian Colin. *Bristol.*
READ, Brian Alfred. *Carshalton, Surrey.*
RUTTER, Neil John. *Manchester.*
TYLER, Brian, B.Sc.(Eng.). *Cambridge.*
VENKATASUBRAMANIAN, K., B.Sc., M.A., D.I.I.Sc. *Tiruchirapalli, S. India.*
WHITFIELD, Robert Stephen. *Walton, Surrey.*

Transfer from Student to Associate

RIVKIN, Philip. *Ponteland, Northumberland.*

Direct Election to Graduate

BUDDS, Donald Eric. *Chelmsford, Essex.*
CLAPP, Raymond George. *Edgware, Middlesex.*
EASEY, William Roy. *Elstree, Hertfordshire.*

Transfer from Student to Graduate

ELLARBY, Colin Leslie. *Hull, Yorkshire.*
KING, Noel Robert Bruce. *Sale, Cheshire.*
O'DONOVAN, Michael Valentine. *Cambridge.*

STUDENTSHIP REGISTRATIONS

The following students were registered on the 26th June.

BAYNE, Robert H. *Nova Scotia, Canada.*
BIRCH, Andrew Nicholas. *London, N.6.*
CARR-GLYNN, Keith A. *Thetford, Norfolk.*
COPLAND, James. *Chippenham, Wiltshire.*
DE-SILVA, Sobanahandi. *London, W.2.*
DUNLOP, Wellington. *Stevenage, Hertfordshire.*
EDDO, Richard Olubuni. *London, N.5.*
EDWARDS, Malcolm H. *Woodford Green, Essex.*
JACOBS, Jonathan W. *London, N.W.6.*
KUMARAVEL, P. *Johore, Malaya.*
LARKIN, Peter A. *Northland, New Zealand.*
LEBBY, Ernest Molu. *London, N.7.*
OKENIYI, Emmanuel O. *Lagos, Nigeria.*
OLEAH, Christopher S. C. *Lagos, Nigeria.*
PARKER, Frank W. *North Harrow, Middlesex.*
PARKIN, John Graham. *London, S.E.19.*
PRABHU, Rama C. V. K., B.Sc. *Cochin, India.*
RAGHVANI, Ramji G. *London, W.1.*
ROBERTS, Michael David. *Cambridge.*
SHAH, Kul Bhushan Kumar. *Bombay.*
SULTAN, Abdul. *Gilgit Agency, Pakistan.*
THORNHILL, Denzil Robert. *Birmingham.*
WATSON, Kevlin J. *Wellington, New Zealand.*

*Reinstatement

U.D.C. 656.052 : 621.396.933.2 : 621.396.96

Some Recent Developments in Marine Navigational Aids

By

J. W. NICHOLS, B.Sc.(Eng.)†

AND

A. C. MacKELLAR‡

Presented at a meeting of the Radar and Navigational Aids Group in London on 18th October 1961.

Summary: The paper describes recent developments in the operation of the older types of navigational aids, lights, fog signals and radio beacons. A novel system of remote control is described and means for automatic operation of remote light stations discussed. Development of a system to give automatic indication of locally reduced visibility is described. A method of employing arc discharge lamps as navigation lights is mentioned. A new system for coding radio beacons is described. A policy for electrical power supplies at remote places is outlined, mention being made of the possible use of solar energy.

1. Introduction

Development of navigational aids in recent years has been devoted to a large extent to work on radar and position fixing systems, such as Decca, Loran, VOR, DME, etc. The rapid development of these systems has been the subject of many varied technical reports to such an extent that changes introduced in more conventional navigational aids have received but scanty attention. It is the purpose of this paper to show what work has been carried out to improve these existing systems.

At first sight it might appear that accurate and reliable position fixing equipment, coupled with the use of radar, would render obsolete aids such as lighthouses, fog signals, radio beacons and buoys. This in fact has been far from the case, the use of the new devices having served to indicate where the older types of aid have been deficient. However, the wide use of electronic techniques has indicated many ways in which the earlier types of aids may be made more reliable and less dependent on human supervision.

It will be obvious that to secure the last objective a comprehensive remote control system had first to be considered. Such a system must be both economical, in installed equipment and in land-line or radio links; unlike most other systems of this type, it need only be capable of a limited speed of operation. The philosophy which most readily meets these needs is one in which all operations at remote stations are automatic. Failure of outstation equipment would automatically introduce some stand-by system, at the same time signalling this action to the control station. It will be clear, therefore, that this philosophy sets certain operating conditions on the items of individual plant making up the navigational aids and power supplies to serve them. It will equally demand the maximum use of efficient electrical equipment as replacement for large uneconomical mechanical items. How this has, in fact, taken place will be discussed in the body of the paper.

2. Remote Control

With a modernization programme in hand for the lighthouse and light vessel service it seems reasonable that the equipment which is installed should be designed for automatic operation to enable the complement of staff at each station to be reduced; eventually it is hoped to reach the stage when stations will be unmanned. It was thus considered desirable to develop a remote monitor and control system which would indicate back to a central control point any change of state of the plant at the remote station, facilities being provided to control the plant operation from the central point should the automatic operation fail.

The Trinity House remote control system therefore has been designed to provide an indication of the following:

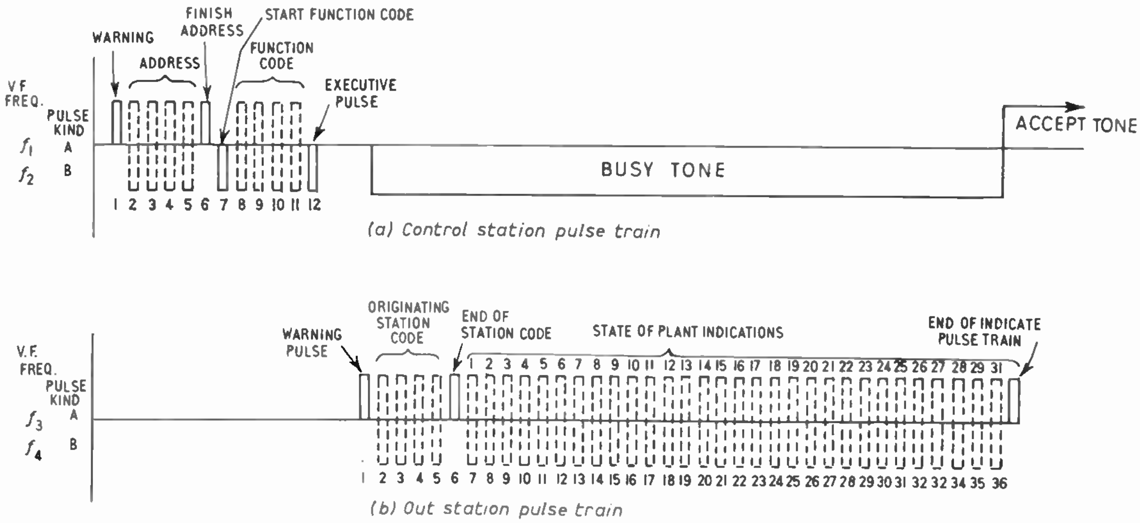
- The identity of the outstation calling.
- Condition of the plant at that outstation.
- Previous state of the plant at the outstation.
- The particular cause for the origination of the change of state signal.

The system will operate over a telephone line or radio link, protection being provided to reject any mutilated code which may be caused by interference or fading.

A total of 16 outstations can be controlled by the central control station; 16 control functions and 30 indications are available at each outstation.

† Formerly Corporation of Trinity House; now with the U.K. Atomic Energy Authority.

‡ Corporation of Trinity House, Research and Development Section, Trinity House, Tower Hill, London, E.C.3.



(a) Control station pulse train

(b) Out station pulse train

Fig. 1. Coding for remote control and supervision.

2.1. *Technique*

The circuits which have been developed for use at the control station have employed cold cathode tubes because of the higher degree of reliability of this device when compared with conventional electromagnetic relays. In addition, it was felt that certain economies could be made as it would be possible to use the cold cathode tube for indication as well as a store of information. The control station has therefore been designed using cold cathode tubes for the coding and switching circuits, these same tubes being used, where appropriate, for visual indication on the control desk. Generation of the v.f. tones used for signalling is by means of transistor oscillator circuits.

As the control of the outstation plant required the operation of a relay contact at the output of the remote control link, it was decided to employ relay circuitry throughout the outstation equipment, the total number of relays required being comparatively small. The equipment operates directly from the 50 V station battery thus giving protection against mains failure and consequent loss of control.

A prototype model of this equipment has been in service controlling a fog signal station for a period of a year or more and much useful information has been obtained both as regards the techniques and the method of presentation and operation.

2.2. *Coding Arrangements for Control and Supervision*

All 16 stations controlled from a single point discussed above can be signalled over a common Tariff D telephone circuit and/or common T100 frequency radio link. Since telephone circuits will be in many cases repeatered it is convenient to carry out all signalling operations by the employment of v.f. tones. To provide for simple separation of "go" and "return" signals the "go" signals from the control station are sent in two low-frequency tones while all outstations make use of a similar pair of tones at a higher frequency.

The sequence of operation may best be appreciated from Fig. 1. The control instruction to any outstation is fully defined by 12 pulses as follows:

- *Pulse 1* of v.f. f_1 is used as a warning signal.
- *Pulses 2–5* may be at f_1 or f_2 and comprise a binary code address system capable of calling any of 16 stations.
- *Pulse 6* of v.f. f_1 indicates termination of address.
- *Pulse 7* of v.f. f_2 indicates preparation for control instruction.
- *Pulses 8–11* are at f_1 or f_2 and define one of 16 central functions in a similar manner to the address.
- *Pulse 12* of v.f. f_2 is used as a total count check.

On the assumption that this control function has caused a change of state of plant at the outstation a similar sequence of tones in this case f_3 and f_4 is started in the reverse direction as follows.

- *Pulse 1* of v.f. f_3 acts as a warning pulse and immediately on arrival at the control station indicates a steady v.f. tone (f_1) which persists until the complete outstation signal has been decoded and accepted by the operator. This tone acts as a busy signal to prevent other outstations seizing the common highway.

Pulses 2–5 as above indicate which of 16 stations is calling and v.f. tones at f_3 or f_4.

Pulse 6 at v.f. f_3 indicates completion of station identity code.

Pulses 7–31 at v.f. f_3 or f_4 indicates the condition of 25 items of plant.

Pulse 32 at v.f. f_3 is used to indicate that the total number of pulses is correct.

Figure 2 shows in simplified block diagram form the circuit arrangements employed to encode and decode the signals discussed above. The diagram is in four parts: the control station encoder, the sending end decoder, the outstation decoder and the outstation encoder.

3. Navigational Lights

It has been the tendency in modern lighthouse installations to employ, whenever possible, electric filamentary type lamps as the light source for the main navigational light. The lamp is operated continuously and the required character is obtained by rotating a rather complicated optical system of reflecting and refracting prisms around the lamp. The capital cost of such an installation is high and the total power required to operate the lamp and rotate the optic is usually several kilowatts. There has been little opportunity of effecting economies in this field as the beam divergence of the system is a function of the source size and the focal length of the lens. Only by reducing the size of the source and by increasing its brightness could a reduction in the overall size of the optical system be made. During recent years there has been an increasing interest by lamp manufacturers in the development of new light sources of high brightness and small physical size for use in the cinema industry. The outcome of this development work has been the compact source 2 kW xenon arc discharge lamp which has a conversion efficiency of 35 lumen/watt, a figure which is twice that of a filament type lamp. This lamp has been fitted into a 500 mm drum lens to provide the main navigational light at Dungeness Lighthouse, the light character being obtained by switching the lamp on for the required flash duration of 1·3 seconds every 10 seconds. This method of operation has a rather serious effect on the life of the lamp, but it has now been possible by careful design of the ancillary circuits to obtain a useful life of about 1300 hours and it is hoped to exceed 2000 hours in the near future. (Useful life includes on and off periods.) The xenon arc discharge lamp operates from a low voltage battery of about 40 volts, although a high voltage pulse of some 50 kV is required to initiate the arc.

Experimental work has also been carried out on the use of a continuously burning 500 watt xenon arc discharge lamp in a very small rotating optical system,

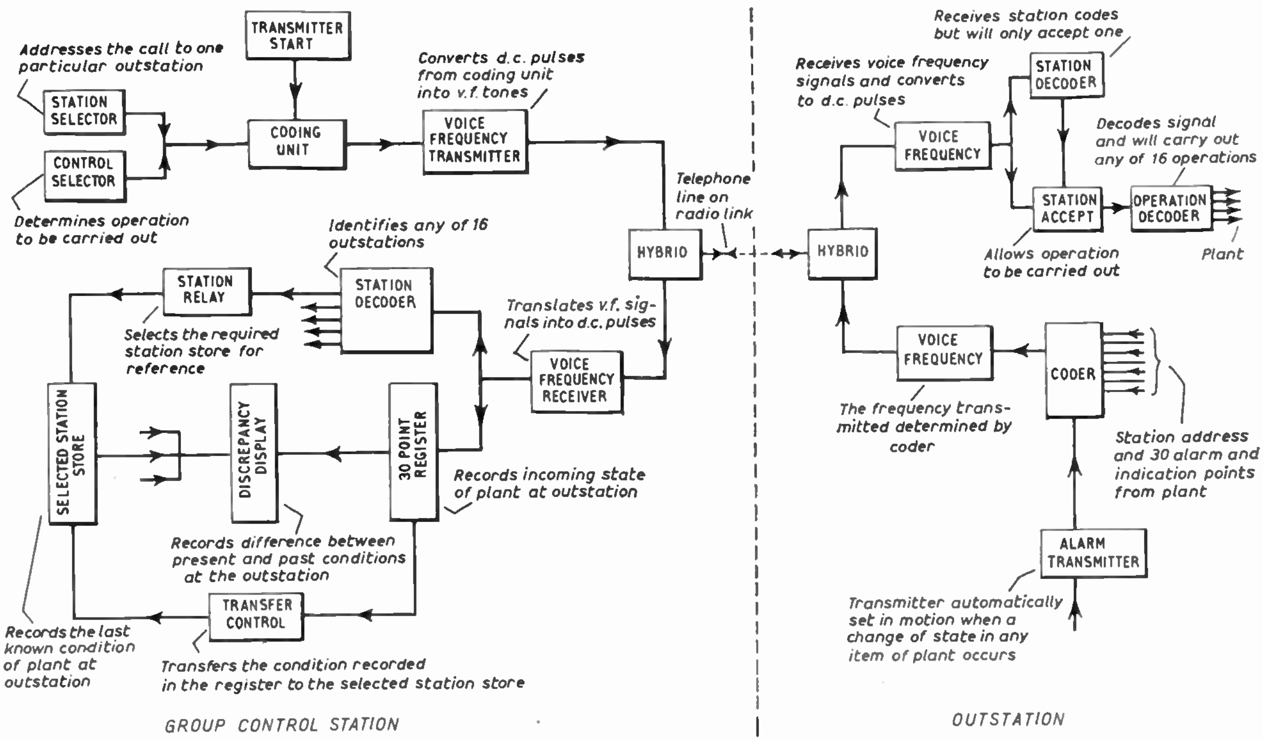


Fig. 2. Block diagram of remote control system.

the focal length of the lenses being only $3\frac{3}{4}$ in; by this means it is possible to obtain a light-to-dark duty cycle of 1 : 12·5 with a beam intensity in excess of 500 000 candela. The cost of the lenses used in this installation is only a few shillings each and in consequence would enable a great saving to be made on the capital cost of the installation.

Development effort has also been directed to the use of semiconductor switching and coding coupled with the use of saturable reactors to provide installations which are devoid of all moving parts. It is hoped to install such a system shortly when a considerable saving in maintenance and a much higher degree of reliability should be achieved.

Table 1 provides a comparison of the costs of the three types.

Table 1

Type of Light	Candle Power	Capital Cost	Running Costs
3rd order 1·5 kW filament lamp	800 000	£6000	£58 p.a.
2 kW xenon in ground and polished drum lens	300 000	£3500	£250 p.a.
500 watt xenon in rotating lens	500 000	£1000	£43 p.a.

Running costs include lamp replacement and energy consumed from the mains at 1·00 pence per kilowatt hour. It should however be remembered that energy costs can be much higher where local diesel generation is necessary.

4. Fog Signalling Equipment

Little change in fog signalling equipment took place in this country between the wars or since. The principles remained the same, namely the compression of large volumes of air to relatively low pressures and its exhaust in bursts through chopping devices, piston, reed or siren. Large and complex horn arrangements were added to attempt to match the impedances and provide some directional properties. It is of interest to note that Lord Rayleigh's earlier work on sound involved many experiments on behalf of Trinity House to improve the performance of some of the earlier systems. Such systems are clearly of very low efficiency and in pursuing a policy of providing equipment capable of remote control some simpler and cheaper device was called for.

Earlier work had, however, led to a general acceptance of fog signals having a note of characteristic low frequency and this does not lend itself to efficient electrical generation. Furthermore, the prime mover in the earlier systems could be run continuously by compressing large volumes of air into reservoir air storage vessels and releasing air for short bursts. An analogous electrical device only exists in the storage battery.

To overcome the first difficulty a three-frequency note was employed in which the interval between tones corresponded to the fundamental frequency of the earlier air driven devices of around 180 c/s. Reproduction then became possible using commercial public address loudspeakers, a typical fully weatherproofed device being available to handle 100 watts. Horn loading was provided by concrete horns.

A multiplicity of such speakers and horns was arranged so that a narrow vertical beam was obtained and at the same time a wide horizontal spread. A

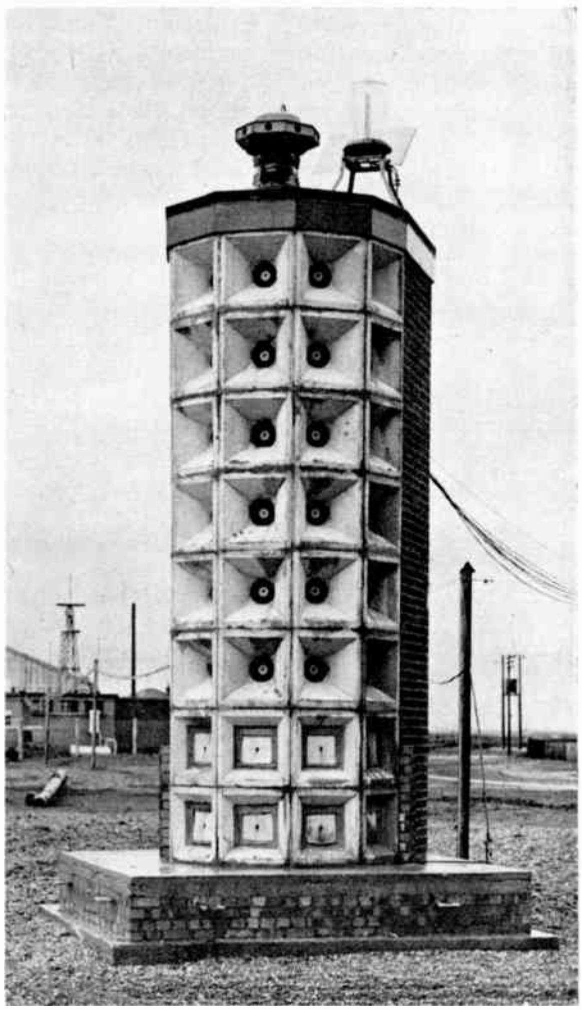


Fig. 3. Experimental multiple loudspeaker fog signal.

photograph of the experimental arrangement is shown in Fig. 3 and the resultant radiation diagrams in Fig. 4. It will be seen that the vertical stacking is equivalent to a line source while the horizontal arrangement approximates to a curved line source. This reduction in beam width accounts considerably for an improvement in performance. The concrete horn structure lends itself to inclusion in a civil design as shown in the new lighthouse at Dungeness (Fig. 5).

The experimental model made use of 30 units of 100 watts each fed from a common three-frequency source derived from three induction generators driven by a squirrel cage motor. Such a system demands the full sounding power to be available and although easy to control and code remotely, does not

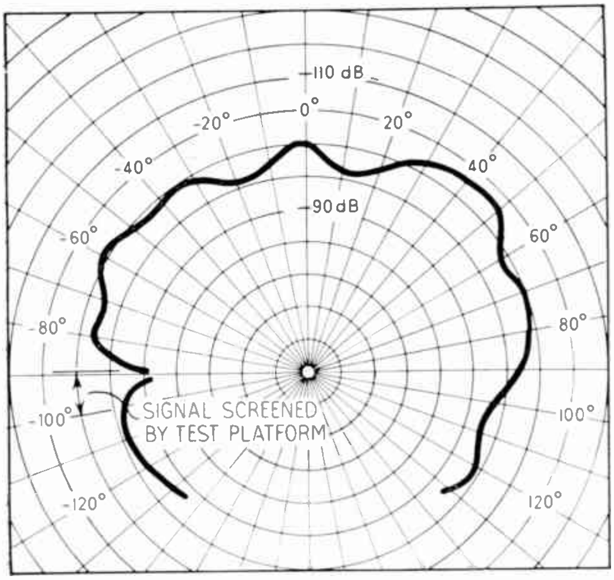


Fig. 4. Radiation pattern of the Trinity House 3-kW electric fog horn taken with microphone at ground level at 100 ft from signal.

fulfil the requirement for power economy. It was felt that where a flywheel inserted between the motor and the generators and the sounding/non-sounding ratio made no more than $\frac{1}{10}$, a peak demand of under 2·5 kW would be adequate to drive this system to provide peak sounding powers of 3 kW.

Development of power transistors, however, indicated a more profitable line of development and work has currently moved in this direction. A pair of OC28 transistors operating from a 24-volt battery was used to drive a single 100-watt unit and an arrangement of four such units is shown in Fig. 6. In this case to ensure the advantage of vertical stacking the four pairs of transistors were driven from a common oscillator. Such a stack can be driven from car batteries

Fig. 5. New Dungeness lighthouse which incorporates electric fog horn and a 2-kW xenon arc discharge light.

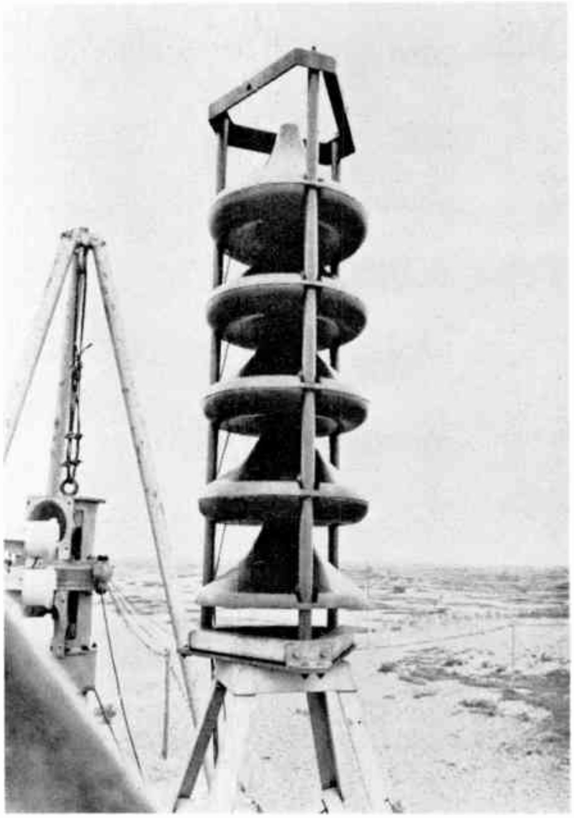


Fig. 6. Assembly of transistor-driven fog horn units.

for about 12 hours without charging on a 1-in-10 duty cycle.

Future development aims at operating at a higher voltage, say 50 volts, and, by using a square wave form of unequal mark/space ratio, to simulate the aural effect of the three-frequency generator system. The satisfactory development of such a fog signal will enable operation from a battery which might be either float-charged, where electricity mains are normally available, or operated on a charge/discharge basis, where the mains cannot be provided. In either case the equipment is intrinsically reliable for a period after the failure of the mains or generator sufficiently long for maintenance action to be taken.

5. Radio Direction-finding Beacon

A network of m.f. radio beacons has been established around the coast of England and Wales to provide direction-finding facilities for the ships using these waters. The beacons are normally grouped in networks of three or six transmitting on a time-sharing basis, each beacon transmitting for a period of one minute every three or six minutes. The transmission from each beacon comprises the station identification signal, which is a two-letter character transmitted in morse code, repeated twice or three times followed by a long mark signal of 25 seconds duration which is the d.f. signal, the transmission terminating in the identification signal repeated once or twice.

It will be appreciated that as the networks are operating on a time-sharing basis a very careful regulation of the transmission time of each beacon is necessary to avoid beacons causing mutual interference. It has been the practice in the past to exercise this control by means of pendulum clocks or chronometers depending on the site, but recent developments in this field have produced an electronic clock coder which not only regulates the time of transmission to a high degree of accuracy, but also enables the transmission character of the beacon to be built up electronically.

A simplified schematic of this equipment is shown in Fig. 7. It consists of a crystal oscillator, a cold cathode divider chain which delivers an output at 4 pulses/second, the basic dot element for the coder unit, with outputs at ½ minute intervals for an impulse clock and 1 minute intervals for beacon switching purposes. The coder unit accepts the basic dot element frequency of 4 pulses/second through gate 1 which is opened by a minute pulse from the timing unit, the output of gate 1 being applied to gates 2 and 3. Simultaneously the minute pulse is applied to the set of ring counters RC1, 2, 3 and 4, the discharge normally resting on RC1. This input pulse steps the discharge round one position to RC2 which opens gate 2 and allows the 4 p/s to be fed to the coding tube which is a ten-way selector tube. The circuit is so arranged to permit the discharge in this tube to rotate up to three times; this combined with suitable switching of the ten output cathodes enables a code letter combination of up to 30 dot elements to be accommodated, suitable outputs being taken to operate the code relay. If, however, the number of dot elements required in the code letters is less than 30, the code reset is brought into operation to reset the coding tube to zero after the required number of dot elements have been counted off. This process is repeated until the required number of code transmissions have been sent as determined by the code repeat circuit. The maximum number of code repeats in any transmission is six, comprising up to four before the d.f. signal with a maximum of two after.

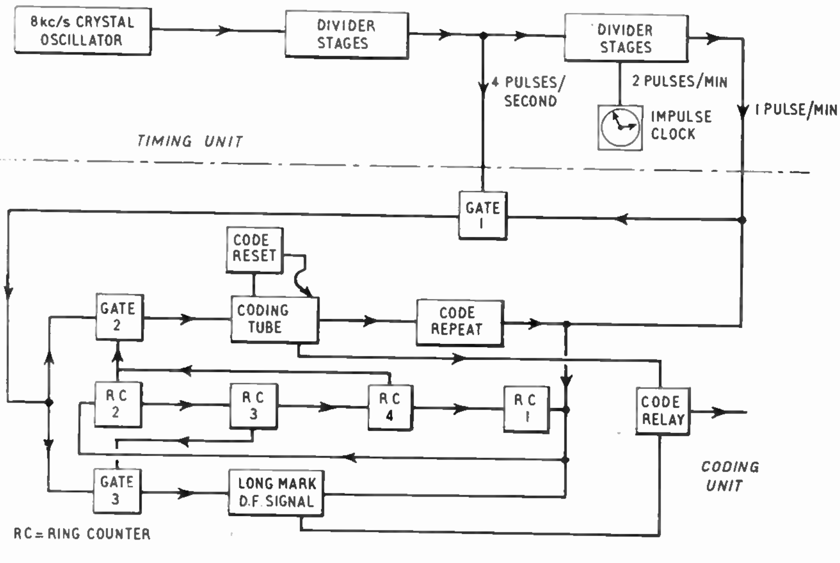


Fig. 7. Simplified block schematic of electronic clock coder.

On completion of the required number of code transmissions prior to the d.f. signal, a pulse is fed to the ring counters, the discharge moving to RC3; this opens gate 3 and closes gate 2. The 4 p/s are now extended to the long dash d.f. signal circuit which switches the code relay to "mark" and proceeds to count off 100 dot elements before moving the code relay to space. A further five dot elements are counted off to give the necessary space between the d.f. signal and the station identification signal which follows it. Yet another pulse is fed to the ring counter to open gate 2 once again and to close gate 3. Dot element pulses are fed to the coding tube which proceeds to transmit the desired station identification letters up to a maximum of two repeats.

When this operation has been completed a final output is applied to the ring counter tubes to move the discharge to RC1, the normal waiting position; gate 2 closes and at the end of the minute period gate 1 closes. The whole equipment is now back in its original state and awaits the next initiating pulse which will start the entire sequence once again.

The entire equipment is operated from the 50 V battery by means of transistor inverters, the station battery being considered as a non-failing power supply.

6. Racon Beacons

Now that a radar set has become a standard fitting on most vessels it has long been considered desirable to provide some form of beacon which would give a characteristic paint on the radar display. Such a radio beacon has been termed a "racon". It is essential that the beacon be received on all radar sets operating within the marine radar band of 9320 to 9500 Mc/s, and it is obvious therefore, that the beacon transmitter must be swept through the radar band at a rate which will ensure that all radar displays receive a paint at fairly frequent intervals. In order to satisfy this requirement the beacon which is described here is swept at a rate of 2 Mc/s by means of a mechanically operated plunger operating on the magnetron cavity.

The beacon is of the interrogated type which remains quiescent until triggered by a radar pulse. It then responds giving a racon-type paint on the p.p.i. It will be noted that a racon beacon gives indication of both range and bearing as opposed to the ramark type of beacon which gives only bearing; the latter is, in general, a free running beacon.

This type of "in-band" beacon does, however, suffer from one important disadvantage as it is not under the direct control of the navigating officer; this fact can be particularly troublesome at close range when the beacon is triggered by the aerial side lobes. In a severe case the entire p.p.i. picture can be obliterated and the navigating officer can take no remedial action. In order to overcome this difficulty a cross-band system has been considered where the beacon transmitter is in the u.h.f. region on a fixed frequency, a suitable converter being installed in the ship's radar set. Under this system the racon facility can be switched out of service by the ship's officer if it is causing excessive interference. Such a cross-band system lends itself very readily to transistorization, the final size and power consumption making it quite suitable for mounting on a buoy where in certain circumstances a racon beacon would be of considerable value. A typical radar presentation of the racon signal is shown in Fig. 8.

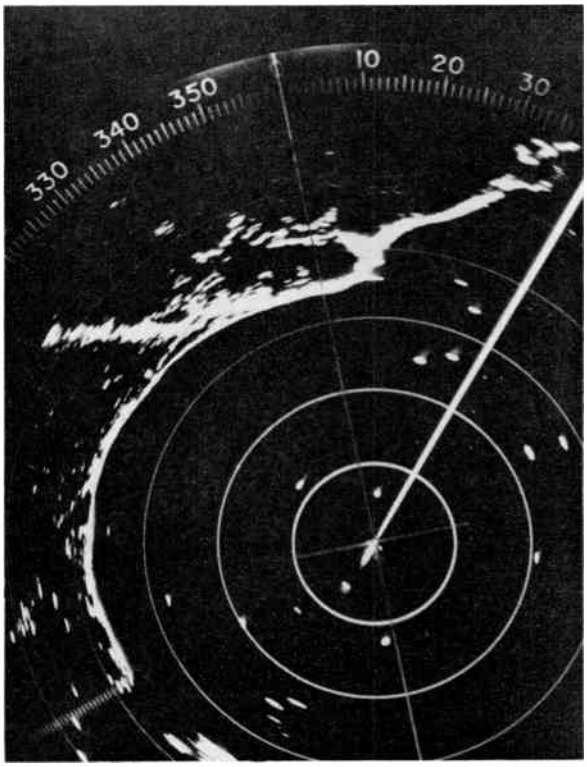


Fig. 8. Radar p.p.i. showing racon signal.

7. Fog Detector

It will be readily appreciated that before any programme of remote control or unmanning of lighthouses can take place it is necessary to design an instrument which will automatically measure the visibility range and initiate the sounding of the fog signal if the visibility drops below a pre-determined level. Such an instrument has been developed by Trinity House and is being brought into service at a number of sites around the coast.

The principle employed by this equipment is to project a beam of modulated light to measure the

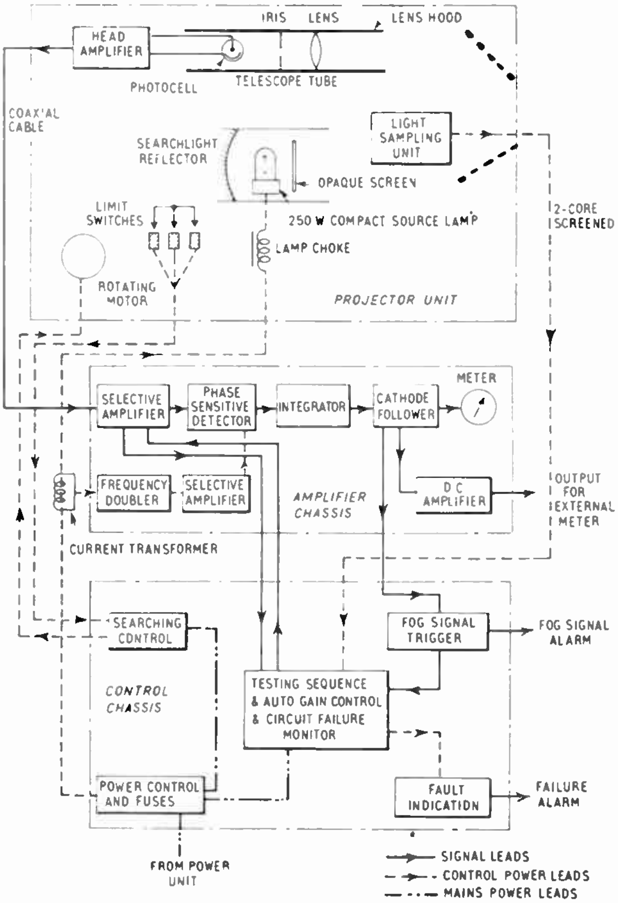


Fig. 9. General arrangement of the fog detector unit.

amount of light which is reflected back from the fog particles in the atmosphere. By using modulated light it is possible to use conventional a.c.-coupled amplifiers in the receiving equipment. Even more important it is possible by this method to eliminate any change in the sensitivity of the equipment because of ambient light levels as the instrument does not respond to what might be loosely called d.c. light, i.e. daylight.

The equipment comprises a powerful searchlight with a beam intensity of about 20 million candelas and a beam divergence of 1 deg. A receiving telescope is mounted vertically above the searchlight with its axis parallel to the horizontal axis of the projector unit; a photocell is mounted just beyond the focus of a 6 in. diameter truvex lens, and an iris is included in the telescope. The projector and receiving telescope is arranged to sweep slowly through an arc of up to 360 deg. On completion of one sweep the assembly sweeps in the reverse direction thus avoiding the use of slip ring connections.

The operation of the whole system is shown in the block diagram of Fig. 9. The photocell is a.c.-coupled to an adjacent amplifier having a low output impedance suitable for connecting to a considerable length of screened cable. A.c. signals at 100 c/s are fed to the main equipment via an attenuator formed by resistances connected to the banks of a uniselector, the resistances being graded to give equal attenuation steps. Signals from the attenuator are fed to an amplifier having an overall gain of 135 dB.

The earlier stages of this amplifier have a twin-T band-stop filter connected in the feedback circuit in order to produce a narrow band-pass amplifier centred on 100 c/s. This amplifier in turn feeds a coherent detector of the Cowan bridge type and subsequently an R-C integrating circuit. It will be clear that the values of R and C chosen will determine the overall bandwidth of the system and thus the signal/noise ratio. It has been found in practice that a time constant of 16 seconds gives completely satisfactory performance and is consistent with an operationally satisfactory rotational search speed for the searchlight. In this context the operational requirement is for a warning if fog subtends an angle of greater than 20 deg of arc, and this gives a search time of 5 minutes so that the total maximum delay from the onset of fog to alarm is 10 minutes, again satisfactory operationally.

The lamp is a high-pressure mercury vapour discharge lamp and may be operated on a.c. or d.c. To obtain modulation a.c. operation is essential but longer life results if one electrode (the cathode) is never allowed to become positive with respect to the anode. This is arranged by the full-wave rectifier circuit shown with the ballast choke in the primary circuit of a transformer. A particular advantage of this method of operation arises since the cathode is the lower electrode and has a characteristic "hot spot" which defines the upper limit of the searchlight beam with considerable precision. A current transformer in the lamp circuit is used to derive the reference waveform for the coherent detector via an amplifier with narrow band characteristics identical to the similar amplifier in the signal circuit. This precaution avoids phase changes between signal and reference waveforms which might otherwise accompany frequency variations.

The output of the integrating circuit drives a cathode follower which feeds both a meter for direct reading of visual range and a Schmidt trigger circuit for automatic initiation of a fog signal.

Deterioration of performance could occur should lamp brilliance decline, lenses and mirrors become dirty or electronic gain fall off. To correct for this at discrete intervals (about once every 10 minutes) a sample of light from the searchlight is injected directly into the telescope by two reflecting surfaces shown dotted in the block diagram. At the same time the capacitor of the integrator is discharged by a contact. The subsequent sequence of operations is shown in Fig. 10 (*a*). The capacitor will now begin to charge and if the gain is correct the amount of light sampled has been adjusted so that the trigger would trip between 12 and 14 seconds. In Fig. 10 (*b*) it has been assumed that the gain is too high and the trip level has been reached after 10 seconds. Under these circumstances the capacitor is short circuited again and the uniselector stepped to reduce the gain by one increment and a second sequence started. The reduced gain leads in this case to operation of the trigger between the 12th and 14th second and the test sequence has been completed correctly. The light reflecting surfaces having been removed, the capacitor is again discharged and allowed to approach the true ambient conditions for 16 seconds prior to restarting a search.

Further development is taking place on a transistorized fog detector operating on the same principle but giving instantaneous coverage of 360 deg. The light source on this occasion is of the xenon discharge type used in stroboscopic photography, the flash tube being mounted in a drum lens. Because of its high conversion efficiency a silicon junction photocell will be used mounted in a drum lens some 3 ft below the light source. The receiver will be a pulse amplifier in this application and will be gated by means of a gating pulse from the flash tube which is arranged to flash once every two seconds. The output pulses are integrated as before and when the output of the integrator reaches a pre-determined amplitude a relay contact will close.

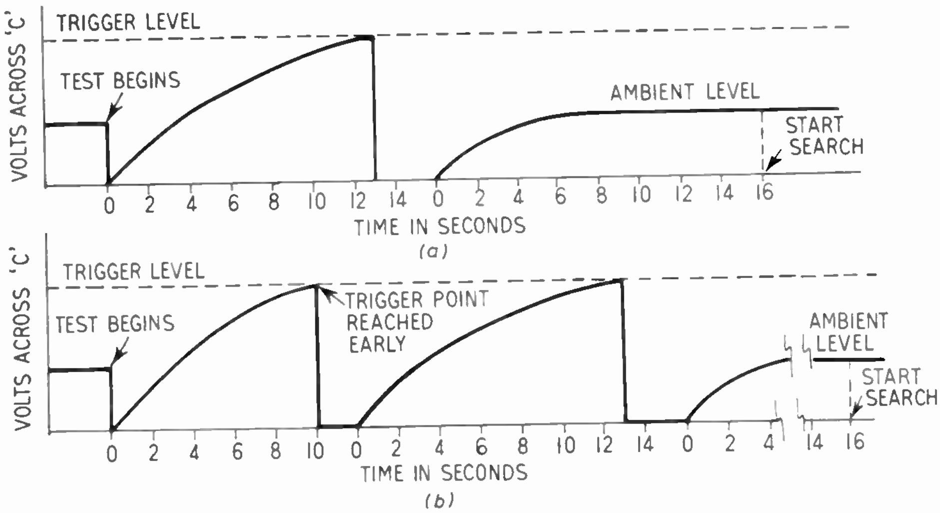


Fig. 10. Test sequence of fog detector.

8. Power Supplies

Prior to proposals for remote control, power supplies to lighthouses could be divided into two distinct groups: those where the electricity grid could conveniently and reliably be connected and others, mainly off-shore lighthouses, where such supply was impracticable. In the former case where personnel were continuously available a diesel generator emergency supply, often d.c., could be manually started in the event of failure of the grid supply. Because of their location away from centres of population lighthouses are more susceptible to interruptions of supply, overhead lines frequently being run in exposed positions on bleak coastlines. Where grid mains could not be supplied it was difficult to justify the use of electricity when paraffin vapour lamps of advanced design were operating reliably except on the basis of eventual remote unattended operation.

In both cases the future of power supplies is linked to the absence of personnel. It has not been regarded as adequate to rely solely on an automatically-started diesel generator operating during mains failure, since the possibility of both failing together is too high. Thus a three-line defence system is employed; firstly in the case of the mains-connected stations the mains is the normal supply, the diesel is stationary and all battery systems fully charged. On the mains failure the diesel would automatically start and connect to the load which in the meantime had operated uninterrupted from the battery supply. It is clear, therefore, that an automatic third line of defence is provided so long as the battery supplies are capable of carrying the load. Restoration of the mains supply will not automatically shut down the diesel generator since experience has shown that a number of temporary reconnections of mains electricity may be expected before permanent restoration of supply. Under such circumstances there is a finite possibility that the diesel starter battery will become discharged. Furthermore, the restoration provides an opportunity for the remote control system to be exercised and for the controlling operator to determine that a temporary restoration may be expected to become permanent.

At off-shore stations or where no mains supply is available it is most economical if diesel generator sets used to provide the supply are operated for short periods at full load rather than for longer periods only slightly loaded. Such operation besides ensuring a higher overall thermal efficiency also increases the operating periods between engine overhauls since full load operation reduces the amount of carbon formed in the cylinder heads.

This method of operation can only be secured by operating from secondary batteries and this requirement has led to the battery operating arrangements for the equipment discussed earlier. Recharging would not be left until the battery was exhausted as again failure of the diesel generator to start might then put the station out of action. In this eventuality the station would operate at reduced power, thus extending the period during which getting a mechanic to site would be regarded as a matter of urgency.

In addition to the maintenance of large light stations, Trinity House is also responsible for more than 200 lighted buoys, which are sited on the coastal waters around England and Wales. The light on these buoys is obtained by burning acetylene gas which is contained in gas cylinders; it is the normal practice to change the cylinders at intervals of approximately twelve months, although sufficient gas is contained in a fully charged set of gas accumulators to last fifteen months. Against this background a number of design studies have been undertaken to provide an economical alternative by providing an electric light with some means of electric power generation on the buoy. The possible methods of power generation include sea water batteries, buoy motion, tidal flow and solar energy; the most promising of these at the present time is the direct conversion of solar energy to electrical energy by the use of silicon photocells which have a conversion efficiency of about 11%.

It will be clear that solar energy is only available during daylight whilst the navigational light is only required at night. Thus a storage battery is essential and it is fortuitous that the silicon cells now available have very suitable characteristics for changing such batteries. It is also obvious that the shorter charging periods in winter are accompanied by longer nights. What may not be so clear is that horizontal mounting yields maximum charge in the summer and vertical mounting maximum during the winter. This latter configuration has two additional advantages: first, fouling by birds is less serious, and secondly, at sea an image of the sun in the sea provides an increase in charge. Vertical mounting can also be arranged so that rotation of the buoy does not materially affect charging.

A standard solar cell available is circular and of 2·54 cm radius yielding an open-circuit e.m.f. of 0·53 volts. Groups of cells in series connections have been used to charge secondary batteries and from curves plotted when the batteries are discharged continuously it has been calculated that 800 of these cells would be sufficient to light a 12 watt buoy lamp when flashing with a duty cycle of 1 in 10. The total area taken by such cells is 1250 in^2 and could easily be accommodated in the type of buoy used.

9. Conclusion and Acknowledgment

We have attempted to show that modern electronic methods can improve the somewhat outdated but in no way unnecessary navigational aids used by mariners around our coast. Electronic techniques, besides providing the communication channels essential to any system of remote control, have also replaced to a large degree the electro-mechanical methods used in earlier control systems and made possible the use of low voltage secondary batteries as emergency power supplies. In this connection perhaps the single device of greatest significance is the power transistor. The replacement of such a device by controlled rectifiers or thyristors will in no way affect the philosophy, but only the circuit arrangements.

A new device capable of unattended operation to indicate reduction in visibility is one novel item which was essential before a remote control system for all navigational aids could be regarded as theoretically feasible.

Capital investment is likely for some time to limit the extent to which these techniques will in fact be applied and further development to improve both efficiency in terms of power requirements and reliability may be expected.

The authors express their thanks to the Elder Brethren of Trinity House for permission to publish this paper.

Manuscript first received by the Institution on 23rd November 1960 and in final form on 23rd June 1961. (Paper No. 743/RNA12).

DISCUSSION

Under the chairmanship of Captain F. J. Wylie, R.N.(Retd.), (Member)

Captain Sir Gerald Curteis†: The introduction of electronics and advanced electrical engineering into lighthouse installations has two ends in view—increased economy and increased efficiency. For instance, at present at a lighthouse station where there is a fog signal, a continuous 24 hour watch has to be kept for fog. The automatic fog detector renders this unnecessary as it will operate the fog signal on the approach of poor visibility—resulting in a considerable saving in manpower. Again, with the introduction of electric fog signals the expense (and weight and space occupied) by the compressing machinery and air-storage tanks—all of which have necessarily to be duplicated, is saved.

With the introduction of the xenon arc, the very heavy expense of a large optic is saved in comparison with the smaller one required for the higher powered light.

Of course the introduction of these inventions must be gradual—not only would it be uneconomical to introduce them into stations before the renewal of the light or machinery is required, but experience has to be gained with their working at one or two stations before proceeding to a fuller programme. This applies particularly to remote control where several stations will be controlled and monitored from one central station. As can well be understood experiments cannot be made with the lights and fog signals in service and therefore all these devices have to be thoroughly tested and proved completely efficient before being introduced.

Mr. A. Harrison (Associate Member): For fog detection the requirement is to detect fog at sea—where the ships are. Even with narrow beams the small vertical separation of the transmitter and detector implies that these beams would not be separated for more than a short distance from the device—perhaps one hundred yards. Beyond this range the behaviour would be analogous to that of a radar set against a target filling the beam, where the received signal falls as the inverse square of the range, causing the fog density measurement to be biased in favour of near fog. Would the authors comment on the desirability of separating the transmitter vertically from the detector, or in the pulsed equipment under development, the possibility of gating the detector to respond only to delayed signals from longer ranges.

In the proposed cross-band radar beacon, the use of a response at a lower frequency, and the poorer directivity of the available aerials at this frequency, would result in an increase in side-lobe responses. Some users' reports had commented on the deterioration of the displayed picture due to side-lobe responses even on the present equipment, where the side-lobe suppression of the marine radar aerial gave about − 30 dB each way. Would the author please comment on this point?

Messrs. J. W. Nichols and A. C. MacKellar (*in reply*): Mr. Harrison's suggestion that the fog detector results are biased in favour of the near fog is in fact true. A theoretical investigation of the problem has shown that in fact the maximum signal return is obtained from a range of 130 feet for the existing fog detector but that appreciable returns are obtained from considerably greater ranges as the curve of signal against range for different visibility ranges does not fall off too sharply. It must be remembered that the fog detector was developed to make measurements of the visible range and to initiate the sounding of the fog signal should the visibility fall below a predetermined value. It has been found operationally that the equipment described in the paper is quite capable of providing this service and in consequence it is not considered desirable to modify the design to enhance its performance at longer range. Most fogs are of a homogeneous nature or of the drifting bank form, both types being capable of detection by the apparatus described above.

This same reasoning is applied to the pulse type of fog detector which is in fact gated but no attempt has been made to use the gating principle to select the sampled volume of space from a greater range as this practice would complicate the design quite considerably due to the higher gains required in the receiver to compensate for the double inverse square law.

It is appreciated that a cross-band racon system operating at a lower frequency would suffer to a greater extent from side-lobe responses than the present in band racon; however, in view of the fact that there would be a continuous presentation appearing on each rotation of the scanner it is felt that it would be much easier and quicker to establish the identity of the racon target by means of the proposed cross-band system. Once the target has been positively identified the navigating officer can remove the racon information from his screen together with any spurious responses which might be experienced at close range due to the poor directivity of the ship-borne racon receiving aerial.

Captain F. J. Wylie (Member): The Merchant Navy is particularly interested in racons and at the moment ships are returning questionnaires recording their operational experience. One of the questions asked was whether the racons cause any serious interference with radar navigation. Of the 119 who have replied, only 12·6% said that there was a navigational hazard and there is no doubt that some of these opinions were formed in the early days when technical defects were present in the operation of the racons.

I have been able to make a personal assessment of the behaviour of the racons on the *Bar* and *Kish* lightvessels during a passage from Liverpool to Dublin as, in matters of this sort, it is so important to understand the degree of hazard which is being reported. I found that at very close range, say half a mile, the *Bar* racon caused obliteration of the screen for two successive revolutions of the scanner in each receiving period. Now, it has to be remembered that the racon transmits only in alternate 5-minute periods and the frequency sweep cycle lasts about 75 seconds. Thus, the ship displays the racon about 8 times every 10 minutes

† Formerly Deputy Master of Trinity House.

and the obliteration could be said to last for about 2 seconds each time. Thus, the total time during which the hazard is present is approximately 16 seconds every 10 minutes.

From the navigator's point of view I do not think that this is a serious hazard; at close range, at night, the flashing light from a lightvessel or lighthouse will obstruct vision over a wide arc for much longer. Navigators do not insist on continuous and instantaneous availability of every navigational aid. It is significant that out of 66 ships reporting on the *Bar* racon only 4 claim that there was a navigational hazard.

The Authors (*in reply*): The questionnaires mentioned by Captain Wylie have been studied and a further analysis made of those vessels which approached to within 3 miles of the racon equipped lightvessels. Under these circumstances the *Bar* Lightvessel indicated that 27% considered a hazard existed and 38·2% had their picture adversely affected and at the *Kish* Lightvessel 35% considered the racon hazardous and 55% suffered interference to their normal radar picture. In addition if a theoretical assessment is made of the time during which it would be possible to experience interference which could interfere with the value of the radar picture the duration is 12 seconds in each period of 75 seconds assuming a receiver pass-band of 10 Mc/s with a skirt slope of 7 dB per Mc/s and a beacon sweep-rate of 2 Mc/s.

In answer to Captain Wylie's last point concerning the availability of navigational aids it is worth noting that there were, in fact, several complaints regarding the time required positively to identify the racon target; an interval of up to 6 minutes could be involved. This figure may not be excessive if the speeds involved are only a few knots, but with the trend to faster vessels such as hydrofoils or hovercraft, a more frequent presentation would be essential.

Major J. G. Cochran (Graduate): With regard to the remote switching of fog warning the criterion here would appear to be to ensure that fog warning devices are switched on under fog conditions at unmanned lighthouses etc. The authors have suggested that this problem should be approached by developing a fog detector that will switch the fog warning on under certain specified conditions of fog. In view of the difficulty of a fog detector discriminating precise fog conditions, would it not be more simple and reliable remotely to switch the fog warning on from the central control post, when fog warning reports indicate that this is advisable?

Have the authors considered the possibility of offering a modification to the ship's radar that would enable the effect of side-lobe triggering to be suppressed? It is suggested that the primary radar should transmit a pulse a few microseconds before the main pulse and from an omni-directional aerial: this pulse would be say 15 dB below the main beam pulse and 15 dB above the side-lobe pulses. This pulse is compared at the racon with pulses received from the main beam and side-lobes, and it is arranged that only pulses greater in magnitude will trigger the transponder.

The advantages of introducing a system of this nature would be that racons could be brought into service without making it mandatory on the shipping lines to introduce any new equipment. Facilities do exist, however, for ships to introduce this modification to their radar or to be built into new radars to completely eliminate the annoyance of side-lobe triggering.

The Authors (*in reply*): Major Cochran's suggestion with regard to switching on the fog signal equipment is not a practical proposition: in the first instance there is no difficulty in detecting fog by electronic means, the Trinity House F.D.1 type equipment performs this function and is in satisfactory operation at a number of sites. In addition the weather conditions which prevail on the coast and in river estuaries are very frequently of a very local nature and it is quite unrealistic to suggest that the fog signal be operated from the central control on information provided by the Meteorological Office. This Service has never claimed a 100% accuracy in predicting weather conditions and could not provide sufficient information to enable the fog signal to be sounded under squally conditions and also when the visibility has reached a fixed pre-determined range.

Consideration has been given to side-lobe suppression on radar aerials but it has not been pursued in view of the fact that modifications would be required to commercial radar equipments. If, however, modifications could be accepted it is then a question of deciding whether it would be preferred from a financial and technical standpoint to modify the entire system as is proposed in the cross-band beacon rather than make modifications to the radar to improve the present in band system. These problems are at present under review and a decision on the future of racon beacons is hoped for in the not too distant future.

Mr. D. E. Nightingale (Graduate): I had assumed during the reading of the paper that the fog detection apparatus was always mounted at the top of the lighthouse and was thus not capable of detecting low-lying fog. The authors have since pointed out, however, that the apparatus would normally be mounted at the "bridge" height of the average ships.

It has occurred to me that if this is so, there must be a blind sector caused by the inability of the detector to see through the lighthouse when mounted on the outside and below the top. Perhaps the authors would care to comment on this point.

The Authors (*in reply*): In reply to Mr. Nightingale it can be stated that the fact that there is a blind sector caused by the tower obstructing the fog detector is quite acceptable under service conditions. It must be remembered that most fogs are of a homogeneous nature and seldom, if ever, occupy a very small volume of space in a fixed position as would be required if the fog detector were to fail to pick up fog in the blind sector. In addition most sites only require fog to be detected over a restricted arc; it is only the isolated rock towers where a 360 deg coverage would be desirable but experience has shown that this is by no means essential.

Mr. J. Hely-Hutchinson: In view of the low transmission loss of sound in water, has anyone explored the possibility of hydrophonic navigational aids?

The Authors (*in reply*): This is an interesting question and the short answer is in fact yes; but because of the very marked reluctance of shipowners to install the necessary receiving equipment on board ship the few underwater sound signals brought into service earlier this century have been discontinued, as very few ships have had the necessary equipment to receive this type of signal.

Mr. C. N. W. Reece (Member): I understand from the authors that for the fog signal, a frequency of 200 c/s is desirable, but for subjective reasons it is best generated at 500, 700 and 900 c/s allowing the rectification action of the ear to supply the missing 200 c/s. Following on from this point, may I ask whether the following possible advantages are taken into consideration?

(a) The ability to generate more audio power at these frequencies using standard amplifiers and loudspeakers.

(b) The shorter horn necessary for effective acoustical loading.

(c) The narrower beam width for a given size of horn.

Also what was the effective audio beam width and was phase relationship taken into consideration? Were the three frequencies chosen to be optimum for the subjective effect allowing for the greater attenuation of higher frequencies in the atmosphere?

The Authors (*in reply*): It can be confirmed that the choice of frequencies for the electric fog signal were influenced by the factors mentioned by Mr. Reece. The beam width obtained for six units stacked vertically is 40 deg to the 6 dB down points at 500 c/s, and naturally the phase relationship is taken into account in order that the advantage of vertical stacking is fully exploited in reducing the vertical divergence of the sound beam and concentrating the signal in the direction it is required. The experimental electric fog signal employed frequencies of 390 c/s, 500 c/s and 600 c/s which are a little lower than those quoted in the question and should be propagated rather better than the higher frequencies previously mentioned; the frequencies selected are not necessarily the optimum, but were taken as a reasonable compromise between the optimum for the ear response and optimum having regard to the propagation in the atmosphere. Further work requires to be undertaken on this problem to find the optimum type of modulation for a fog signal to ensure the best results not only regarding range but also of what might be termed "punch-through" or the ability of the signal to make itself detected in the presence of ship noise, wind noises and wave noises.

The Chairman: I think that this has been an exceptionally interesting paper and it has evoked a lively discussion. I am not sure that the authors brought out the significant fact that the benefit of the research carried out by Trinity House is felt by all the Lighthouse Authorities of Great Britain and Ireland. Further, the aids to navigation which come to be established as a result of this work are international, in that they are of course equally available to ships of all countries. This is not only a good thing but it also makes it necessary to remember that a navigational aid which requires the ship to fit special equipment before she can avail herself of it is a matter which obviously involves international considerations.

Finally, I would like to express the thanks of all of us to Mr. Nichols and Mr. MacKellar for an excellent paper.

DEVELOPMENTS IN MARINE ELECTRONICS

The Navigation of Hovercraft

At a recent meeting of the Institute of Navigation it was stated that the operational requirements and navigational equipment of hovercraft will depend on the type of craft and the route on which it is employed. For the immersed-sidewall hovercraft, which is not expected to travel over 30 knots, conventional navigation gear fitted in present river vessels will be adequate.

For the amphibious air-curtain hovercraft, the main new operational requirements will be for navigational equipment for operation over beaches and mud flats, and for use at speeds higher than 20 knots, for which this type of hovercraft is designed. The general opinion was that hoverways on the beaches and over the sea might be necessary in the interests of safety.

Radio navigational equipment for track keeping in hoverways was discussed and the equipment which was outlined varied from position-fixing aids such as marine radar, Decca Navigator and direction finding equipment, to modern dead reckoning systems with compass references and Doppler navigational radars. For anti-collision purposes, true motion radar carried by the hovercraft was proposed. Closed-circuit television to improve lookout astern was also discussed.

New Receiver for the Royal Navy

Bulk production of units of a new ship-borne receiver for the Royal Navy is now commencing at the Bracknell factory of Racal Electronics. A major design feature of the receiver is its ability to operate in close proximity to high power transmitters. A specially designed r.f. tuning unit is incorporated not only to reduce interference from such transmitters to a minimum (100 dB attenuation is achieved on signals 5% off-tune) but also the circuits have been designed to be sufficiently robust to avoid burn-out should the receiver be tuned accidently to an adjacent transmitter.

The tuning of the receiver is performed by a precision decade tuning unit which enables the receiver to be left tuned accurately to its channel over very long periods. The receiver has a frequency range of 1–30 Mc/s and has facilities for operation on d.s.b., s.s.b. and c.w. The units of the receiver are mounted on a robust cabinet which incorporates features such as anti-condensation heater, CO_2 fire extinguisher inlet, and a perspex faced door to render the installation splash-proof. Provision is made for forced-draught cooling which can be connected to a ship's air trunking system.

A Shop-window for the Electronics Industry

Since it was first held in 1957, the "Instruments, Electronics and Automation" Exhibition has expanded to such an extent that this year at Olympia, London, 550 firms took part and an attendance of over 90 000 was recorded—a 20% increase over the last exhibition in 1960. The proportion of overseas visitors increased from 1 in 19 to 1 in 15 and the rôle of the exhibition as a place for doing business is reported to have shown striking results. From the viewpoint of the engineer wishing to keep abreast of the latest developments, there was much to attract his attention—possibly *too* much for those with wide interests—and there seemed to be a general desire on the part of exhibitors to present the technical features of their products informatively. Valves and semi-conductors as well as passive components customarily receive adequate backing in this respect, but measuring instruments have not always been described in detail. The applications of electronics form an important feature and demonstrations of complete systems intended for widely varying industries were prominent.

Elliott-Automation demonstrated the computer system that is now being installed at the Spencer Steel Works of Richard, Thomas & Baldwin in South Wales, and which will result in the nearest approach to full automation yet achieved. The system is termed a "hierarchy of computers", all of which are in communication with each other. The computers are organized in the same general pattern as a human chain of command, accepting orders, parcelling out the work to other computers, making sure that the machines in the works are used to the best advantage and finally applying physical control to such equipment as rolling mills. (A paper describing the system was read at the Brit.I.R.E. Symposium on Industrial Electronics recently and will shortly be published in the *Journal.*)

Ekco Electronics showed new nucleonic instruments for measuring the thickness and profile of flowing sheets of material on a production line. One of these continuously measures the thickness of coating—paint, plastics or tin—being applied to a product. The uncoated material is measured before it goes through the process and the measurement is fed to a memory unit. A second measurement is made at the end of the process and the two are compared. The difference between the two thicknesses represents the quantity of coating applied and can be used as a basis of quality control.

In the field of miniaturization, Mullard have produced a "condensed" assembly of components, in less-than-matchbox size, made up of 34 transistors, 90 diodes, 162 resistors and 64 capacitors. It is a 15-stage shift register for computing applications and is *forty times* smaller than the equivalent device produced in conventional printed-circuit form. The same company shows examples of a technique of depositing circuit elements on to a very thin sheet of glass, developed for the Ministry of Aviation. The glass substrate is only one thirty-second of an inch thick, yet in an area about a quarter that of a visiting card, 4 transistors, 5 diodes and 12 resistors have been "imprinted". These glass "stamps" can be mounted one on top of the other to produce extremely compact assemblies.

McMichael Radio have produced sub-miniature modules for the British *Black Knight* rocket, which have a parts density of 15 000 components to the cubic foot, but they announce that new techniques have much improved on this figure, and a dice-sized unit containing 6 transistors, 8 resistors and 4 capacitors can be produced, giving a parts density of 250 000 per cubic foot, while still meeting the rigorous performance specifications of the original units.

One of the exhibits by Ferranti was data-logging equipment which can be hired on a monthly basis by companies requiring facilities for collecting large quantities of data on physical variables for statistical or mathematical studies of industrial processes. Logging equipment is normally mounted in a central cabinet inside a caravan for mobility. Variations in physical characteristics of the process can be sampled at speeds as high as 1650 values per minute or as low as 1 value per second.

Generators for use with telecommunications systems were shown by Marconi Instruments. The double pulse generator provides jitter-free single or double pulses variable over a wide range of amplitude, duration and repetition frequency. It has comprehensive trigger and sync facilities, together with precise control of pulse position. Also shown was a nanosecond pulse generator which is a discharge-line type of pulse generator giving positive or negative pulses of very short rise-time, accurately controlled stable amplitude, and negligible sag or overshoot. Pulse duration, determined by lengths of internal coaxial cable, is variable from 2·5 to 100 ns (0·0025 to 0·1 μs) in five steps.

In the control and instrumentation field Ultra Electronics showed a servo-system for reactor temperature control in a nuclear power station. After referring to the demanded and actual temperatures, the servo adjusts the position of control rods which regulate the activity within the reactor core and thereby adjusts and maintains the reactor power output level in accordance with the demand. Magnetic amplifiers are employed to ensure the high order of reliability and stability which is required. (This equipment was the subject of a paper read at the Brit.I.R.E. Symposium on Industrial Electronics and will be published in the *Journal* in due course.)

The slab thickness meter designed by English Electric Aviation enables the thickness of red-hot slabs of steel to be measured as the ingot passes through the rolling mills. It is designed to measure thicknesses of from 3 in. to 20 in. and consists of a scanning unit and a display unit. When the slab appears within the field of view of a telescope on the scanning unit, a pneumatic actuator causes a mirror to move vertically upwards. As it does so, the measuring head sends a train of electrical pulses to a counting circuit. A photo-cell detects the radiation from the ingot reflected by the mirror and the pulses are switched off when the sightline passes above the top of the slab. The counting circuit is coupled to a logic unit which energizes an illuminated lamp display showing slab height in inches and quarter inches.