THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 28 AUGUST 1964 NUMBER 2

NATIONAL ELECTRONICS RESEARCH COUNCIL

NEARLY two years ago the Institution's Research Committee commenced a survey of radio and electronics research in Great Britain.† The ensuing report and recommendations were submitted to Admiral of the Fleet the Earl Mountbatten of Burma, K.G., then President of the Institution. His subsequent discussions with appropriate Government Ministers, and heads of Universities, Colleges of Advanced Technology, and industry, resulted in all these interests agreeing to form the National Electronics Research Council.

N.E.R.C. became a legal entity on 20th July 1964 by incorporation under the Companies Act 1948 (Certificate No. 813236) and was also granted a Licence by the Board of Trade pursuant to Section 19 (1) of the Companies Act.

This event was marked by a press conference held in the Lecture Room of the Institution on Wednesday, 22nd July 1964. Lord Mountbatten, Chairman of N.E.R.C., presided, and was supported by Mr. L. H. Bedford (a Vice-President of the Institution), Mr. Michael Clark, Sir Robert Cockburn, Mr. O. W. Humphreys, Sir Joseph Lockwood, Sir Harry Melville, Sir Gordon Radley, and Mr. C. O. Stanley. Messages welcoming the formation of the National Electronics Research Council were read from the Minister of Aviation, the Rt. Hon. Julian Amery, M.P., and the Secretary of State for Education and Science, the Rt. Hon. Quintin Hogg, Q.C., M.P.

A pamphlet issued by the Council stated that:—

"N.E.R.C. will initiate and encourage the co-ordination of electronics research in industry, universities and colleges of advanced technology, and government research establishments. N.E.R.C. is *not* a research association, but a central co-ordinating body which will indicate gaps in research, show where additional effort is required, suggest priorities, and how to prevent unnecessary duplication of research.

† *The Radio and Electronic Engineer*, **25**, No. 1, p. 3, January 1963.

Government responsibility

"The government has a particular responsibility for research for defence projects, but in the process is often able to inject new ideas into industry which are suitable for commercial exploitation.

"University research is often of a speculative nature but it is necessary to tackle the problem of integration of such research with an overall policy so that immediately there is some result from speculative research, ideas can be adopted and exploited by industry.

"In this connection it must be remembered that many companies invest a great deal of their own money in subjects of their own choice for research and development, but industry shares with universities the problem of how to tackle projects which are often too extensive for the resources of an individual company, particularly where the immediate sales return may not provide adequate recompense for development effort.

Co-ordination of effort

"A discussion of this sort of problem between government, industry and universities would aim at the co-ordination of effort which would be of benefit to the country as a whole and not detrimental to the interests of any single manufacturing organisation. Such co-ordination would still permit industrial organisations to exploit commercially the 'know-how' derived from their contribution to a co-ordinated project. Industrial firms would be more willing to pool their results if the research is co-ordinated in its early stages, rather than if collaboration is attempted when considerable work has already been done from their own resources.

"Such co-ordination requires understanding between top level management in the manufacturing industry, chief scientists of government departments controlling research establishments and of user ministries including the Defence Services, and representatives of the universities and colleges

of advanced technology. This will now be facilitated by the constitution of N.E.R.C.

Other selective research projects

"Working Parties representative of the interests already mentioned are already examining how best collaborative research may be carried out in Great Britain.

"N.E.R.C. recognises, however, that in such a vastly expanding field, it is impossible to cover each and every aspect. Restoration of the balance of initiative in favour of British industry demands a highly selective direction of research projects. This is fundamental to the whole concept of N.E.R.C., which is perhaps in a unique position to try to deal with the problem.

"The Council is conscious that security may sometimes prevent publication and utilisation of much defence-sponsored work. It is hoped that frequent reviews by the Government of the security of their projects, followed by discussions in N.E.R.C., may enable more information, previously on the secret list, to be released to industry and universities.

Assisting universities and colleges

"The place of universities in the national effort can be helped by N.E.R.C. in resolving two main problems:—

(*a*) To secure closer understanding between universities and industry. The responsibility of universities to preserve freedom to carry out their researches with detachment is recognised but it is also realised that both universities and industry can benefit from a much closer understanding of each other's aims and difficulties. N.E.R.C. hopes to encourage any steps to improve the sharing of knowledge and experience between universities, industry and government research laboratories.

(*b*) To harness university effort on projects which are too extensive for the resources of a single university. A solution may be the collective handling of major projects by a group of universities.

Recruiting brains

"The adequate recruitment of staff for both pure and applied electronics research is another vital problem which N.E.R.C. is actively considering. A review is being made of all the scholarships and awards for electronics research which are available in Great Britain through a multiplicity of agencies. N.E.R.C. will provide those who seek scholarships with complete information of all available bursaries and is also willing to assist the granting bodies in the selection of candidates.

"The Council will continually investigate the possibility of obtaining additional grants for electronics research.

Retrieval of information

"If information on new work, new developments, new techniques and new equipment is not disseminated to the research worker, and if he cannot easily and speedily retrieve information on a problem facing him, he cannot work to the best effect and will be in constant danger of duplicating work that has already been done elsewhere.

"Ensuring exchange of knowledge has now become a gigantic task because every year over one million scientific papers are published throughout the world circulated through some 20 000 journals. In electronics alone, some 25 000 papers and documents were issued in 1963 and this 'information explosion' must be controlled.

"N.E.R.C. has already started within its own organisation to tackle this problem. It is obviously one which no individual company is likely to tackle. A Working Party comprising representatives from government, industry and the universities has therefore begun its investigations with a view to providing a Central Information Service for the research worker in the electronics field."

The Council's pamphlet also refers to the work done by the Institution before N.E.R.C. was incorporated. Under the heading "Setting the Pace" it is stated that:—

"Even before the Council was incorporated, it was obvious that the first task would be to survey the present research effort because this information was not available in easily accessible form. With the ready collaboration of The Institution of Electronic and Radio Engineers (which acted as our secretariat), the universities, colleges of advanced technology, government research establishments, and industrial organisations, an interim report was laid before Ministers and industry in October 1962. The final report was circulated to government departments, industry and universities in April 1963."†

N.E.R.C. is controlled by a General Committee consisting of representatives of Government Departments, the Committee of the Vice-Chancellors and Principals of the Universities of the United Kingdom, the Conference of the Electronics Industry, and the Society of British Aerospace Companies.

The Secretary of the Institution, Mr. Graham D. Clifford, has also been appointed Secretary of N.E.R.C. and he is assisted by Mr. T. M. Aitchison as Technical Officer.

† *Proc. Brit. I.R.E.* 1, No. 5, pp. 113–41, July–August 1963.

U.D.C. 621.395.455

Multi-channel Open-wire Carrier Telephone System—Salisbury (Southern Rhodesia) to Kitwe (Northern Rhodesia)

By

A. MUNRO (*Associate Member*), †

G. H. MACKENZIE (*Associate*)‡

AND

C. W. M. ANDERSON

(*Associate Member*)‡

Summary: The use of open wires to provide a high-capacity carrier telephone system in conditions where radio interference is very light and climate extremes unknown is described. The system is specially applicable to underdeveloped territories where two pairs of relatively light construction may be used to provide at least 132 channels over distances of 500 miles.

1. Introduction

In 1954 the possibility was recognized of utilizing a very much higher frequency range on open-wire lines for a multi-channel carrier telephone system because of the special circumstances existing in the large and geographically remote Federation of Rhodesia and Nyasaland. It was accepted that in more developed countries with neighbours at a similar stage of advancement there would be almost complete use of the radio spectrum between 150 kc/s and 550 kc/s by a multitude of aeronautical and maritime navigational aids, many of very high power, against interference from which it would be impossible to construct and maintain an open-wire carrier line. Such use of this radio spectrum in the Federation is limited and radio-frequency power is not very great from any of the navigational aids now functioning. Furthermore decreasing use of such aids in favour of v.h.f. systems is anticipated.

In the design of a suitable line configuration it was apparent that the ideal approach of constructing a completely new pole route carrying only the pair or pairs of wires for the wideband project could not be considered. The problem then was to add to or modify existing open-wire routes to accommodate at least 120 channels. Usually the type of route encountered has at least two pairs of 0·104 in copper-covered steel wire so constructed and transposed that one twelve- and one twenty-four-channel system is carried. There are many other pairs transposed to carry four-channel carrier systems and magneto or automatic party-line circuits in the voice range with frequent 'taps' to subscribers. With an eye to economics there could have been added to the route, with reasonable cost, two additional pairs which would have provided a total of six groups of twelve channels each. Unfortunately, something more than this capacity was needed to tide over the time before a move to wide-band techniques on either radio or cable could be justified.

2. Theoretical Approach to the Problem

The use of the higher frequency range up to 300 kc/s to accommodate 24 channels on one pair of open wires is now well known and is achieved by spacing the pairs so used by some four feet whilst at the same time reducing the wire spacing to six inches. Generally, therefore, only two pairs of such capacity are found on a single open-wire route. Unfortunately because of the co-ordination necessary with the twelve-channel system carried by the same pair in the frequency range 36 to 150 kc/s it is seldom that more than three or at the most four repeater sections can be tolerated, making a total distance of some 250 miles. Clearly a more radical approach is necessary when twice as many channels at double the distance are required without a greater expenditure on open wire construction. There are several limiting factors in the exploitation of open-wire routes in either wide-band or heavy-loading techniques, namely:

(1) With realistic wire-to-wire and pair-to-pair spacings the increase of both near-end and far-end crosstalk§ with increasing frequency seriously limits the use of more than two systems simultaneously using the band above 150 kc/s.

(2) Even with systems below 150 kc/s very good geometry must be maintained in the open-wire route, together with frequent transpositions to make the use of more than four 12-channel systems practicable over long distances.

† Formerly Chief Engineer, Ministry of Posts, Federation of Rhodesia and Nyasaland.

‡ Ministry of Posts, Southern Rhodesia.

§ Crosstalk between pairs on a telephone route represents a combination of several components due to electric and magnetic couplings by way of different paths. The crosstalk propagated in the disturbed pair in the opposite direction to the wave in the disturbing pair is 'near-end' and the component propagated in the same direction is 'far-end'.

After consideration of these factors and recognizing the necessity to minimize capital outlay on an open-wire route which ultimately will become purely 'audio' in character the following approach was made:

(3) Accept the fact that only one system in the spectrum above 30 kc/s would be placed on the route.

(4) Recognize that for the major length of the route, i.e. 550 miles (890 km), the entire spectrum from 6 to 552 kc/s should be available for through circuits if necessary.

(5) Shorter sections up to 80 miles (128 km) should carry, for junction purposes, two further 'super-groups' in the range 564 to 1052 kc/s.

The concept outlined in (3), (4) and (5) is essentially one of recognizing the limitations in respect of crosstalk attenuation between pairs on an open-wire route and therefore not attempting to use more than one transmission system. Naturally this one system must provide as many channels as possible and hence the frequency range used must be as high as practicable without resorting to uneconomic equalization of the attenuation/frequency characteristic. Requirement (4) defines the number of channels capable of being used for conversations from one end of the route to another as 132 when allowance is made for limitations in filtering. Finally (5) calls for another 120 channels in successive short sections, of up to 80 miles each, in the route.

Now in order to transmit frequencies of up to 1052 kc/s without difficulties arising from 'absorptions'[1] due to neighbouring pairs and also to insure reasonably low noise levels frequent transpositions are necessary. Absorptions occur when the crosstalk from a pair into the route as a whole suffers a 180 deg phase shift before being re-induced into the pair under consideration. Since the attenuation of the pair between transposition points is negligible it is permissible to equate,

$$x \propto \log_e\left(\cot \frac{2\pi f}{v} d\right)$$

where

x = near-end crosstalk

f = frequency in c/s

v = velocity = 300×10^6 metres/second

d = distance between transposition points in metres

x will tend to $-\infty$ when

$$\left(\cot \frac{2\pi f}{v} d\right) = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2} \text{ etc.}$$

Obviously the first point, at $\frac{\pi}{2}$, is the one defining the limit for transmission in a continuous frequency band on the pair.

Equating for d to ascertain the maximum distance between transpositions:

$$d = \frac{\pi}{2} \cdot \frac{v}{2\pi f} = \frac{v}{4f}$$

Expressing this in km and kc/s for convenience:

$$d(\text{km}) = \frac{75}{f(\text{kc/s})}$$

If 1052 kc/s is the bandwidth to be transmitted it would be necessary to test with a frequency about 25% higher to insure a smooth characteristic. In practice about 17·5 per km or 28 per mile would be acceptable. However, as standard spacings are 40 poles per mile it is evident that a transposition on every pole becomes necessary.

Having accepted that there will be only one system on the route, since it would be impracticable to control far-end crosstalk sufficiently with normal pole setting deviations in the older type of open-wire route to accommodate a second system in the spectrum above 300 kc/s, it is necessary to assess the maximum usability attainable. Clearly a go-and-return two-pair mode of operation is realistic if a limit to bandwidth is imposed as in this case by crosstalk absorption. Furthermore the controlling factor becomes near-end crosstalk which appears as 'echo' if, as is essential, the two directions of transmission occupy the same frequency spectrum. Such operation is of course implicit in the standard C.C.I.T.T. method of carrier channel derivation.

To determine repeater intervals and operating levels for such a system it is convenient to start with an assumed tolerable level of 'echo' for the longest link likely to be set up—in this case 900 km. The C.C.I.T.T. gives guidance[2] in respect of near-end crosstalk between the two directions of transmission of a telephone circuit, and quotes the minimum ratio as 35 dB. It would be as well at this stage to assess some of the factors which in a link of this nature determine the location of repeater stations before proceeding to the actual calculation involved. Firstly, it is essential that all major and minor towns are 'dropping' points for one or more groups of circuits and it follows that these should also be repeater points preferably containing gain regulating equipment. Secondly, it is convenient from the viewpoint of level standardization to make the intervals between each of these 'dropping repeaters' equal, although due to variations in distances it is very unlikely that similar spacings will exist in different parts of the route. Naturally, there will be circumstances where unequal spacings of repeaters in a 'dropping' section will arise, notably where power supplies or buildings already exist or access roads make for convenient maintenance.

Now theoretical consideration of the magnitude of near-end crosstalk attenuation between horizontally spaced pairs indicates that with centre-to-centre spacing of between 70 and 78 inches and a wire-to-wire spacing of from 8 to 4 inches the amount is likely to lie between 63 and 76 dB. This will be true only for a route carrying the pairs in isolation and the calculated values are shown on Fig. 1. In fact on a route carrying many other pairs there will be tertiary couplings which may reduce the crosstalk attenuations to between 50 and 65 dB for the spacings mentioned when measured at 550 kc/s.

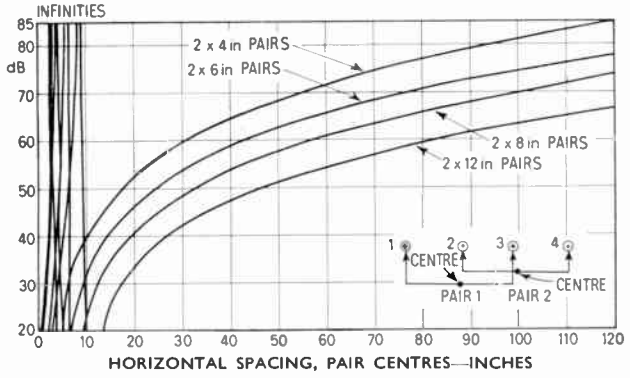


Fig. 1. Near-end crosstalk attenuation versus horizontal spacing for 4-in, 6-in, 8-in and 12-in wire-spaced pairs. 158 lb/mile copper-steel pairs, wire diameter 0·104 in. Impedance: 4 in, 521Ω; 6 in, 571Ω; 8 in, 604Ω; 12 in, 653Ω. Pair-to-pair crosstalk only, no tertiary couplings. Slide rule accuracy, i.e. ± 1 dB.

'Infinities' refer to the 'interleaving' of pairs to achieve zero coupling. In practice however it is impossible to maintain the spacing accuracy to achieve such inductionless arrangements. An example is shown in the inset diagram.

By labelling the wires 1, 2, 3 and 4 it can be shown that the condition for zero coupling arises when

$$\frac{1\text{-}4 \times 2\text{-}3}{1\text{-}3 \times 2\text{-}4} = 1$$

One solution of this equation for 12-in wire-to-wire spaced pairs gives 8·485-in centre-to-centre of the pairs. The second solution is unattainable in practice. The 'infinity' for the 12-in pair is thus shown on the graph at about 8 inches.

The original measurements were made using American 'copper-weld' wire but final construction was made with the British product 'copper-ply'. This is made by using an electrolytic deposition method.

Possible repeater station spacings may be tested by assuming an even distribution using between 10 and 40 repeaters and the most pessimistic attenuation, for the standard conductor of 0·104 in copper/steel on 80 pairs of porcelain double-grooved insulators per mile, of 0·775 dB per mile at 550 kc/s with a wire spacing of four inches.

The deterioration arising out of many tandem repeater sections may be expressed as

$$E = x - 10 \log_{10} n - A$$

where

E = 'echo' or the hearing by the talker of his own speech, expressed as an attenuation,

x = short section near-end crosstalk,

n = number of tandem repeater sections,

A = line attenuation per repeater section.

Table 1

10	repeater sections	of	55 miles	of attenuation		42·6	dB/section.	
20	,,	,,	,, 27·5	,,	,,	,,	21·3	,,
30	,,	,,	,, 18·33	,,	,,	,,	14·2	,,
40	,,	,,	,, 13·75	,,	,,	,,	10·65	,,

Substitution of the factors shown in Table 1 gives values of 'echo' lying between −2·6 dB for ten repeater sections using 8-in wire spacings to 38 dB for 40 repeater sections using 4-in wire spacing. Whilst this latter value is only 3 dB above the C.C.I.T.T. recommendation and thus leaves a small margin for crosstalk in equipment, it is based on quite pessimistic estimation of near-end line crosstalk attenuation.

The next step was to make some estimation of the likely signal/noise ratio attainable with available wide-band amplifiers and bearing in mind the requirement dictated by the foregoing 'echo' limits of 40-repeater intervals. An experimental section of 8-in spaced pair with transpositions at every pole had been built in the vicinity of Salisbury and noise measurements on this confirmed that due to the well-balanced construction the only significant interference came from broadcast stations and aviation beacons. In the range below 550 kc/s only the latter were observed. The highest observed level was −63 dBm. It was noted that the use of well-balanced and shielded transformers with longitudinal 'stopping' coils was essential. Whilst the general level of background noise due to 'static' was so low as to be negligible, it was decided to determine operation levels on the basis of experiencing noise/interference levels of −63 dBm on every section of the route. The following relationship indicates the signal/noise ratio S/N dB which follows from the use of channel transmit level to line L dBm, repeater section attenuation A dB, number of repeater sections in tandem n and noise level N dBm per section.

$$S/N = L - 10 \log_{10} n - A - N$$

Assuming that in the few channels that will be affected by interfering signals it will be possible to use

compandors it would be reasonable to make S/N = 40 dB to give an estimated channel overall signal/noise ratio of 60 dB, allowing a 20 dB improvement due to the compandor.

Rearranging the formula and substituting known values we find the required transmitting level L:

$$L = 40+16+10{\cdot}65-63$$
$$= +3{\cdot}65 \text{ dBm, at 550 kc/s.}$$

Accepting that a 'load rating' of about 20 dB will be required for 132 channels it follows that the amplifiers must be able to cope with +24 dBm output level without distortion and a bandwidth of 6–552 kc/s.

To anticipate a little it may be noted that despite early hopes of a transistorized amplifier meeting the above specification being available, it has transpired that a valve-type wide-band coaxial system amplifier has had to be used with consequent difficulties in power supply at many remote repeater stations; however, this aspect will be treated later.

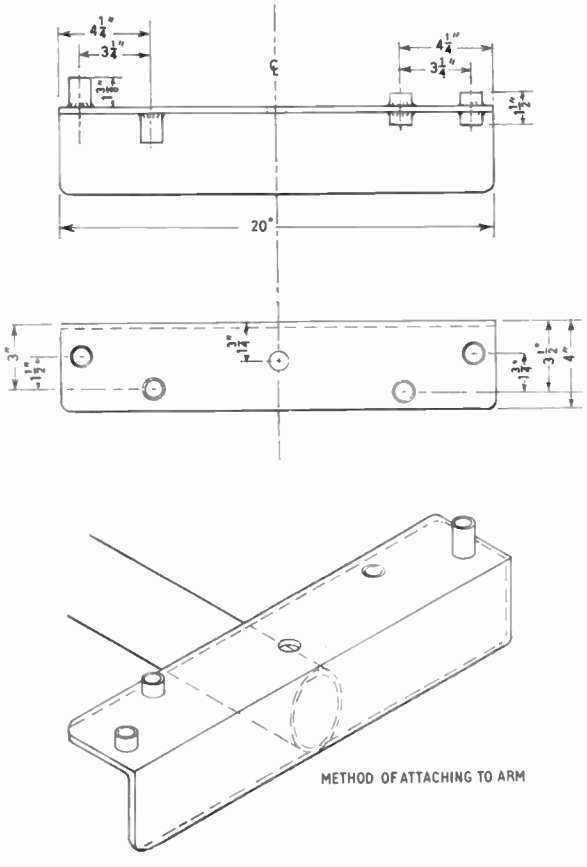
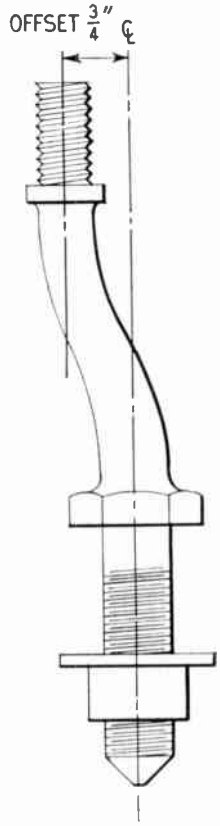


Fig. 2. Special transposition bracket. 4-in spacing high-low, end-arm mounting. It will be seen that at the right-hand end, where the spacers give a 'level' condition to the spindles, the $1\frac{3}{8}$-in spacing added to the insulator groove diameter of about $2\frac{1}{8}$ in and the wire diameter of 0·104 in results in 4-in wire-centre-to-wire-centre separation. At the left however, where a level difference of $1\frac{3}{8}$ in is provided between the wires, it will be found that a wire spacing of 4 in is achieved.

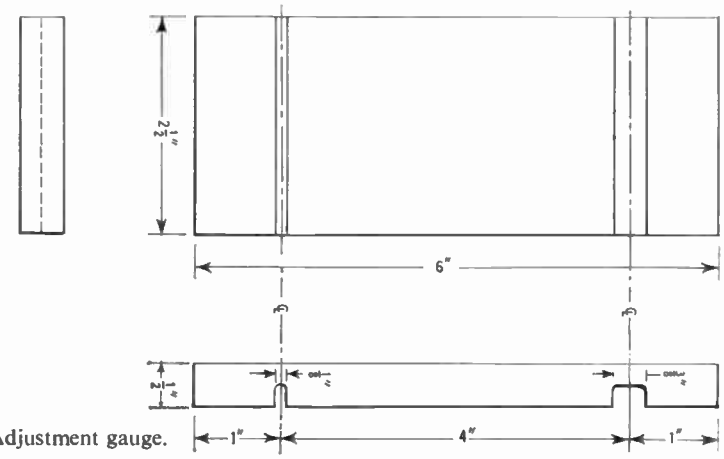
3. Practical Design and Construction of the Open-wire Route

At the outset it was realized that there were many difficulties in achieving a 4-in wire spacing free of contact and twisting troubles. However, there is considerable advantage in having the very equable climate enjoyed by Rhodesia in that no provision need be made for ice loading on open wires, neither do wind strengths exceed 80 miles/hr. It is practicable therefore to specify wire tensions in normal day working temperatures of 160–180 lb or about 15% of the breaking strain of the copper-steel wire used. In these circumstances the geometry of the route may be maintained reasonably easily. A further safeguard against wire-to-wire contacts is achieved by a special design of transposition bracket as illustrated by Fig. 2. It will be seen that the layout in plan of this bracket is essential to attain the 4-in wire-to-wire spacing whilst allowing the use of standard double-groove porcelain insulators. The additional feature of the bracket is variation of wire position in elevation in the span from level at one arm to $1\frac{3}{8}$ in difference at the next arm. On test this proved very effective in eliminating contacts whilst easily providing the transposition on the bracket and maintaining capaci-

Fig. 3. General appearance of the route. The repeater hut and pole-head transformers can be seen in the centre of the picture.



(*a*) Cranked insulator spindle.

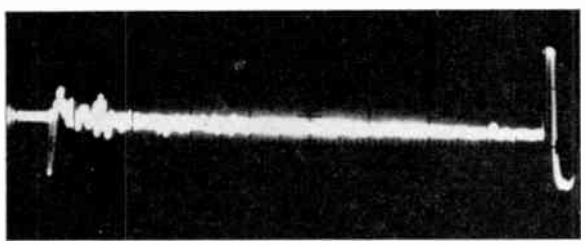


(*b*) Adjustment gauge.

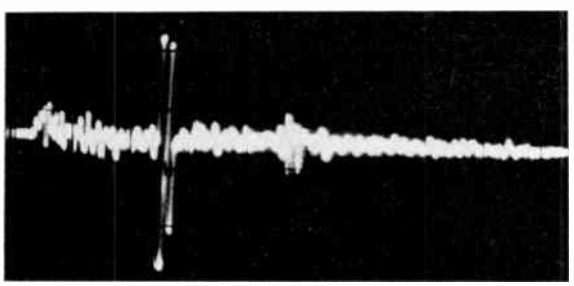
Fig. 4. Device for providing adjustment in spacing.

tance balance of each wire of the pair. The attachment of the bracket to the arm and the general appearance of the route is shown on Fig. 3.

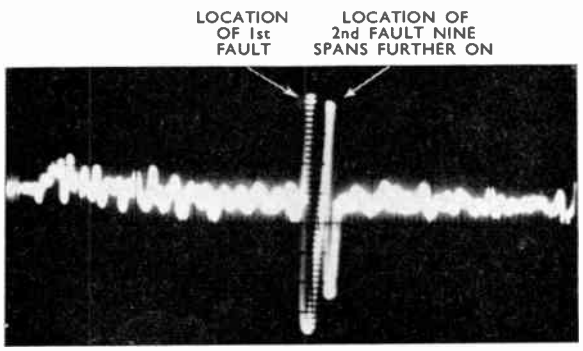
On completion of the first repeater section of approximately fourteen miles the usual tests for impedance, near-end crosstalk, noise and attenuation were carried out, since for a four-wire system these were considered the only significant properties of the pairs. It became apparent as construction and testing progressed that irregularities did arise due to wire spacing inequalities in broken terrain and the additional device, of providing for adjustment in spacing when required, was adopted. The method is shown in Fig. 4(*a*) which is in fact a cranked insulator spindle whilst in Fig. 4(*b*) is shown the gauge with which adjustment to the required accuracy could be set.



(*a*) Line with minor faults only.



(*b*) Line with major and minor faults.

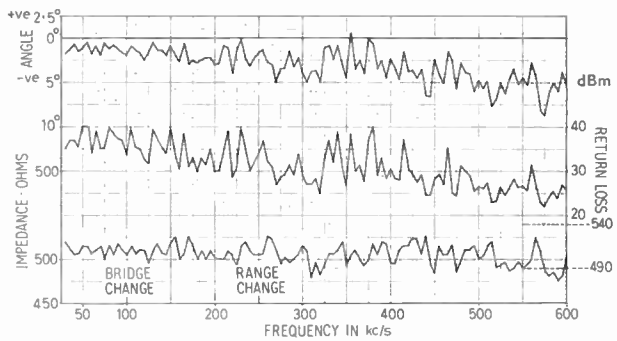


(*c*) Trace (*b*) scale expanded to locate a twisted-wire fault accurately. The twisted wires did not short-circuit the route as oxidation on the wire surface provided a measure of insulation. It was thus possible to 'see' the reflection of the test pulse from both faults. The expansion of about 2 : 1, i.e. time scale halved from trace (*b*) to trace (*c*), is easily seen.

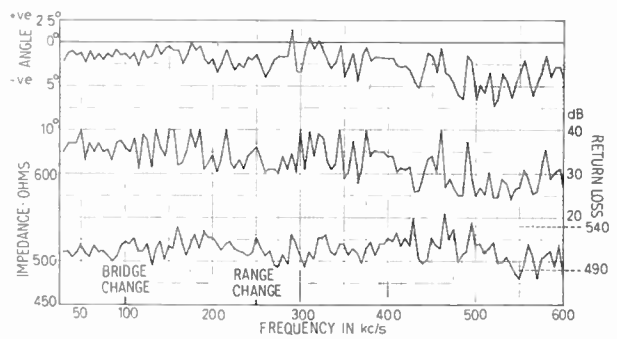
Fig. 5. Irregularities in wire spacing faults shown on typical oscilloscope traces.

4. Testing and Adjustment of the Route

At the outset it was apparent that very much more expeditious location of construction errors and material faults was necessary and to this end a pulse echo fault locator[3,4] was modified for use on a balanced open-wire line. This unit enabled irregularities in line impedance to be located with precision and corrected to meet the arbitrary requirement of a 26 dB return loss. Irregularities discovered by this versatile instrument included high resistance joints, wire spacing deviations of $\pm 0{\cdot}25$ in, sub-standard insulators and other low-insulation faults. Figure 5 shows in (*a*) an acceptable section, (*b*) two points separated by nine spans where construction crews had inadvertently twisted the wires during erection, and (*c*) shows the trace of the faulty condition expanded to make location easier.



(*a*) Pair 1-2.



(*b*) Pair 9-10.

Fig. 6. Graphs of line impedance characteristics of typical section between repeaters no. 7 and 8.

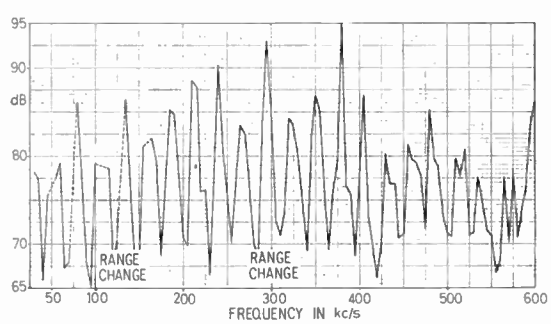


Fig. 7. Near-end crosstalk attenuation between repeaters no. 7 and 8, pair 1-2 and 9-10.

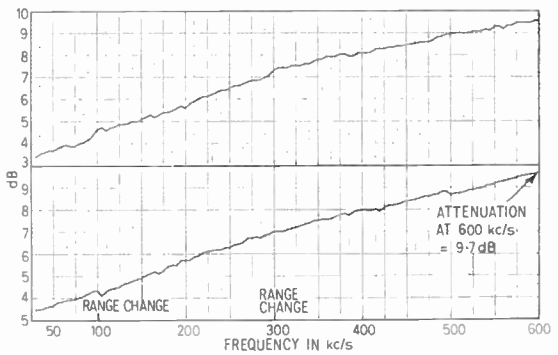
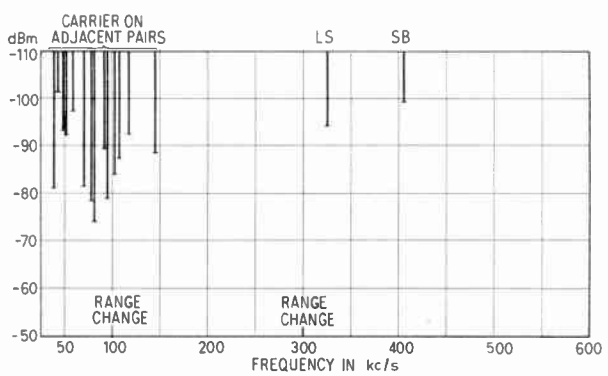
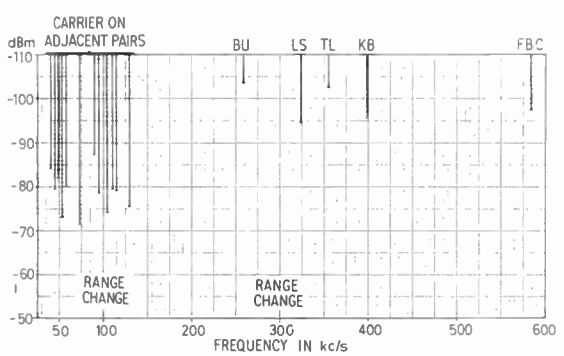


Fig. 8. Line attenuation between repeaters no. 7 and 8, line length 13·44 miles. The upper curve refers to the pair which has become known as HF1 whilst the lower curve is for pair HF2. The route being operated on a 4-wire basis, one pair is the 'go' and the other the 'return'.



(*a*) Pair 1-2.



(*b*) Pair 9-10.

Fig. 9. Noise interference characteristic of the section between repeaters no. 7 and 8.

Measurements of characteristics of a typical section between repeaters no. 7 and 8 are shown in graph form on Figs. 6–9. It will be seen that return loss is at worst 22 dB whilst the impedance approximates to 514 ohms. Near-end crosstalk attenuation is somewhat better in this section than the desired 65 dB. The attenuation shown is for dry weather and includes about 0·5 dB for matching transformers. The wet weather attenuation has been established as 0·775 dB per mile at 552 kc/s. The graphs of noise interference (Fig. 9) indicate that in this section only carrier systems on adjacent pairs induce significant voltages and in any event are to be withdrawn on final commissioning of the h.f. system. The interference from the coded aviation beacons 'LS', 'SB', 'BU', 'TL' and 'KB' can be ignored. All levels shown are in dBm.

5. Construction of the Route

Construction commenced in January 1961 and was completed in September 1961. Figure 10 shows the 550 miles between Salisbury and Kitwe, the two terminals of the system. Figure 11 shows the final route diagram with all details of special lead-in cable and equipment. Mention will be made later of the 'power feeding' system shown on this diagram. The appearance of the open wire route is conveyed by Fig. 12 which also illustrates the type of repeater hut used. The detail of terminating arrangement using 'J' spindles is clear from Fig. 13 from which it can be seen that rearrangement of the spacers allows for either a 'flat' or 'staggered' termination as necessary.

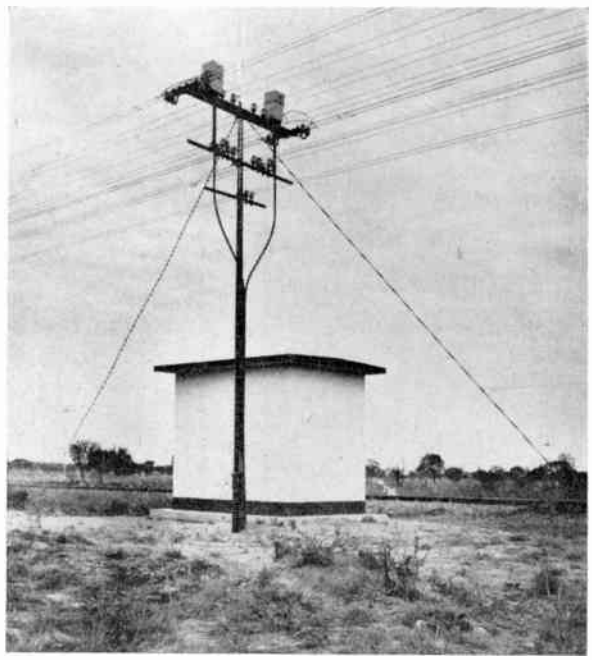


Fig. 12. Repeater hut and pole-head transformers on the open-wire route.

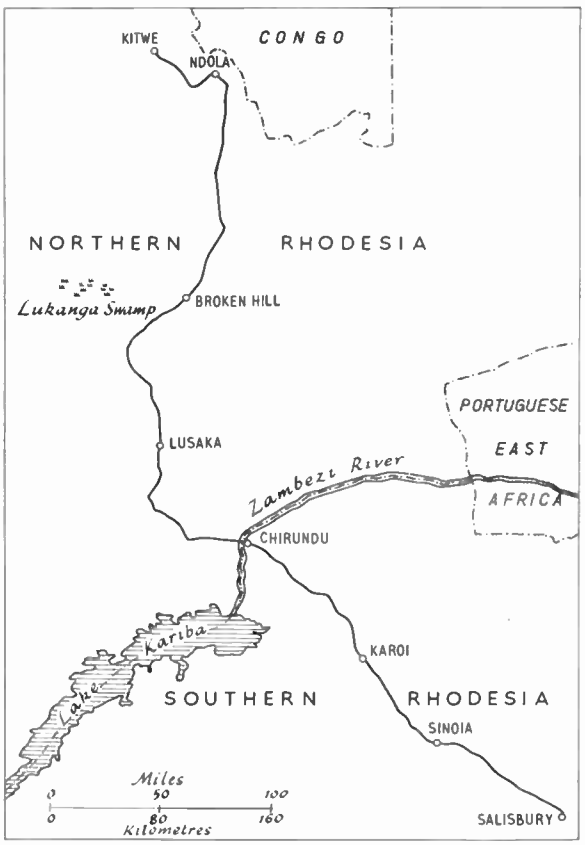


Fig. 10. Route between Salisbury and Kitwe.

Opportunity was taken during construction to deviate some 100 miles of pole route to an alignment along newly built roads in Southern Rhodesia. Unfortunately in Northern Rhodesia pole spacings are much greater and varied from 28 per mile to 20 per mile and in fact, due to absorptions, determine the upper limit of usable frequency.[4,5] However measurements confirm that operation up to 552 kc/s is practicable and in short sections up to forty or fifty miles (Kitwe–Ndola) the spectrum up to 1 Mc/s is available for two further super-groups of 60 channels each giving reasonable telephone quality with 'echo' at 20 dB down. Due to the very short delay in such open-wire transmission 'echo' is not objectionable.

Details of construction practice are given in the following sections.

5.1. *Wire Spacing*

In order to provide minimum crosstalk between pairs the insulator supports are established at the

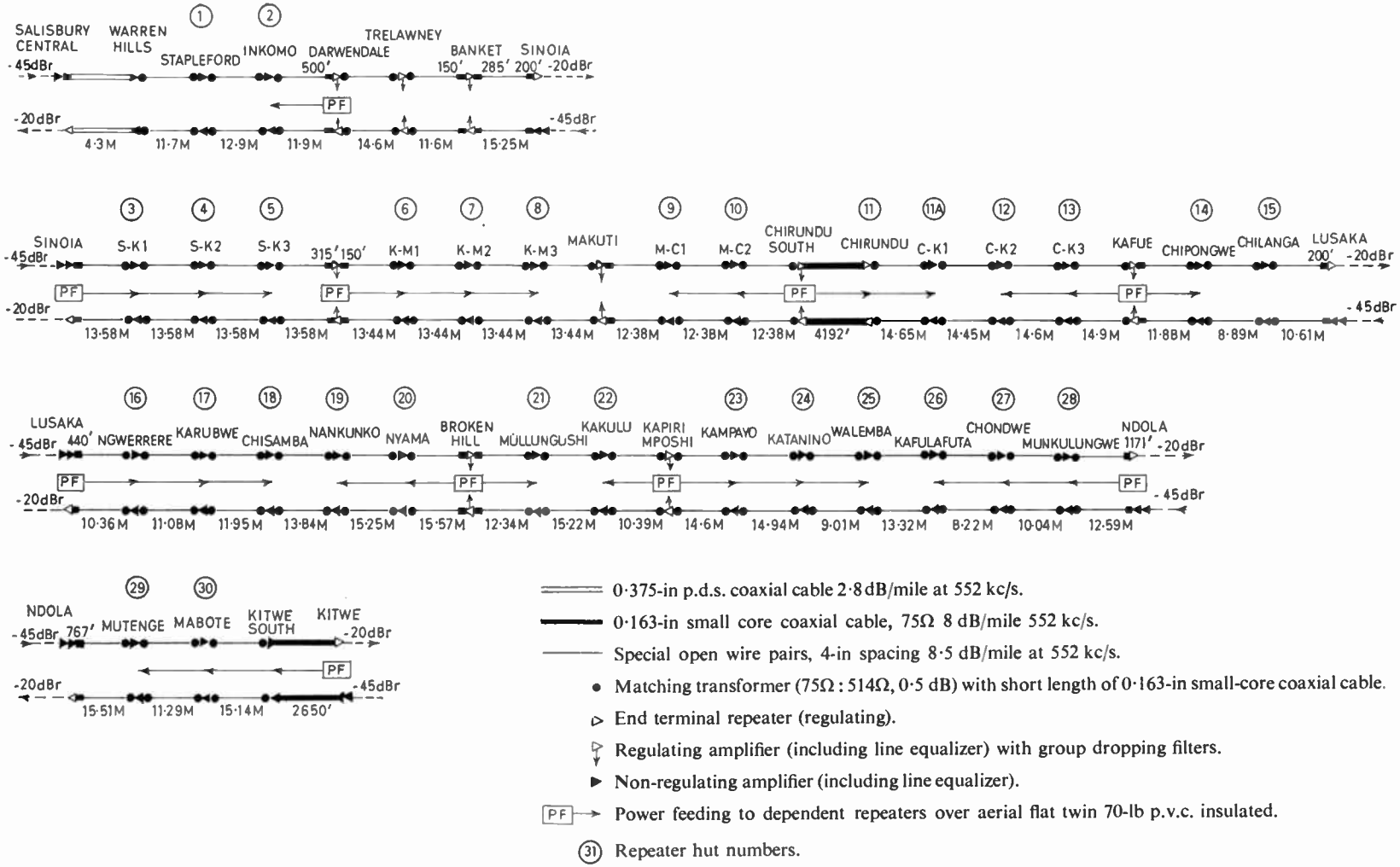


Fig. 11. Route diagram showing the types of cable and positions of repeaters.

The level dBr is in reference to the originating level of the system which in this case, for test purposes, would be 0 dBm at the two-wire channel point. Thus — 20 dBr is equivalent to — 20 dBm. At the stations Sinoia, Lusaka and Ndola full group translating and inserting equipment is fitted and the standard input levels to this equipment are — 20 dBr whilst the output is — 45 dBr, i.e. a loss of 25 dB which is dealt with by the non-regulating and equalizing amplifier shown in each case. The output level to the line at each of the above stations and also at the repeaters, in each direction, is + 3·5 dBr.

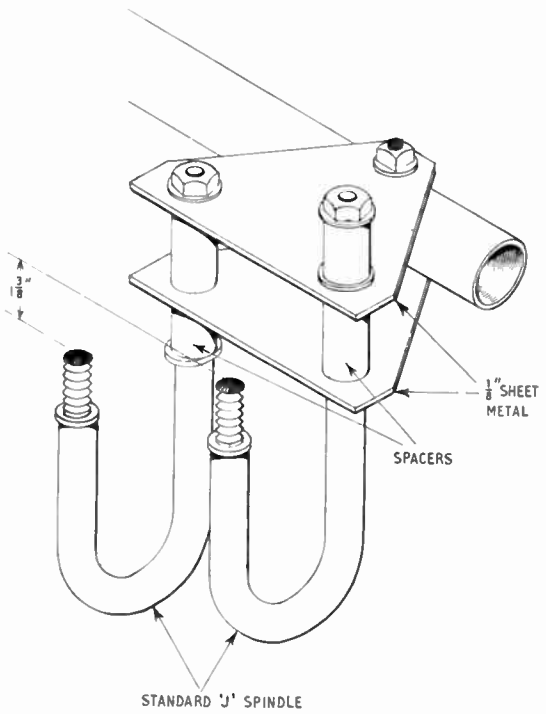


Fig. 13. Terminating bracket using 'J' spindles. The positions of the spacers allows 'flat' or 'staggered' termination.

extreme ends of the 81-in (2·05 m) cross-arm, giving a centre-to-centre pair spacing of 78 in (2·0 m).

5.2. *Conductors*

The conductors are copper-covered steel-core wire 158 lb/mile (44·5 kg/km), 0·104 in (2·64 mm) in diameter, drawn to a tension of 160–180 lb (73–82 kg) depending upon the span length.

5.3. *Insulators and Spindles*

Insulators are large white porcelain double-grooved type T1 manufactured according to the South African Bureau of Standards Specification 161 : 1955 for low-voltage porcelain insulators. The diameter of the groove used is $2\frac{1}{8}$ in.

Insulator spindles are of the standard type as specified in British Standard Specification B.S. 161 : 1949—Telegraph Material.

At angle poles, one or more of the specially cranked type of spindle are fitted to give the necessary adjustment for 4-in wire spacing between centres. The simple adjustment required is illustrated in Fig. 14. After the wire spacing had been checked with the 'go/no-go' gauge the spindles were firmly locked in position. One-eighth inch (3·2 mm) felt washers are used to 'bed' the insulators firmly on the spindles.

5.4. *Ties*

To secure the line wires to the insulators the conductors were first wrapped with a pre-formed, hard-drawn copper-covered steel reinforcing wire of helical shape and then secured with annealed, pliable binding wire. This method had been found from previous experience to be economical and effective, particularly in preventing the thin copper coating of the line conductor from being chafed.

5.5. *Joints*

The line conductors were jointed with compression type sleeves, crimped twice on each half with simple, hand-operated, double-lever compression pliers.

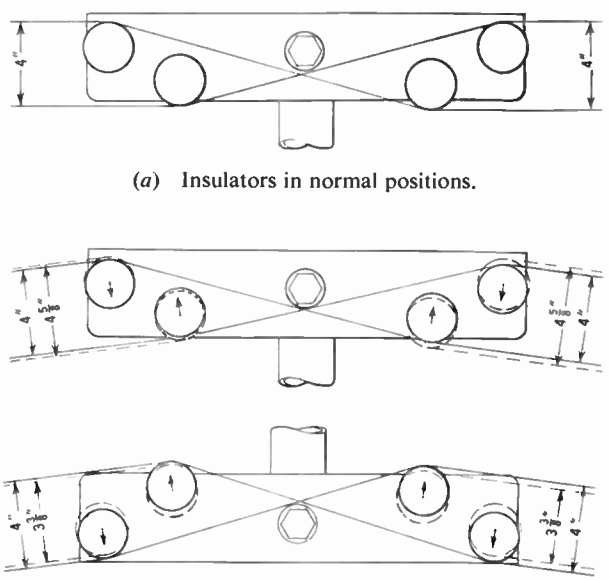
At termination points Britannia type joints were used—cranked compression type sleeves were too expensive.

5.6. *Power Crossings*

To maintain the impedance regularity earthed power guards were built over the bare line conductors in preference to using insulated line wire for the h.f. pairs.

5.7. *Transpositions*

Each pair was transposed at every pole, but to improve far-end crosstalk attenuation one transposition was omitted at mid-section on the 'go' pair and one transposition at the quarter and three-quarter points on the 'return' pair, in each repeater section.



(*a*) Insulators in normal positions.

(*b*) and (*c*) Insulators in new positions when using cranked spindles on bends.

Fig. 14. Use of cranked spindles to offset termination positions.

5.8. *Trees*

In order to derive the maximum shielding effect against lightning as few trees as possible were cut, so leaving a narrow, clear swathe parallel to the route. In certain areas, however, the swathe had to be widened considerably to prevent damage to the line caused by elephants pulling down trees.

5.9. *River Crossings*

To circumvent any problems involved in maintaining the span of 438 metres across the Zambezi River it was found expedient to lay a coaxial cable from Chirundu across the road bridge in a section just less than 1·6 kilometres long.

6. Equipment and Buildings

6.1. *Equipment Installation*

All equipment is of 75 ohm unbalanced configuration and is connected via solid-core coaxial cable to pole-head transformers mounted as shown in Fig. 12. Each enclosure contains the two transformers necessary for the one direction of transmission; very little spacing is necessary because of the very low repeater gain.

Lightning protection is provided by the expedient of designing the line winding of the transformer on the same principle as a 'drainage' coil, i.e. leakage inductance between the two halves not greater than 40 μH and current carrying capacity of not less than 400 A for several-hundred microseconds. The centre point of the line winding is directly connected to a good 'trench' earth (not greater than 3-ohms resistance) at the foot of the terminal pole. Experience to date does not indicate the need for protective gaps from each leg of the open-wire line to earth.

The repeater huts are simple reinforced concrete buildings 6 ft by 8 ft in plan and 8 ft high (1·83 m × 2·44 m × 2·44 m) cast on the site in one piece using sheet-steel and angle-iron shuttering, including the roof slab, in one operation.

6.2. *Power Supplies*

It proved impossible to obtain transistor amplifiers for the unattended and remote repeaters and conventional thermionic-valve coaxial repeaters have had to be used. This has presented a problem in that power feeding of about 120 W per station has had to be provided to as many as three stations spread over forty miles from the main station. It will be appreciated that safety of public and administration staff in both normal and fault conditions of the power-feeding circuit has been paramount.

Power is fed along the route on a twin copper conductor, 70 lb/mile weight (140 lb/mile total) contained in a 'figure-of-eight' p.v.c. heavy sheathing. The feed is longitudinal through the repeater stations to a terminal earth at the last station receiving power on the particular chain. After attempts to use direct-current feeding had failed, due to the conversion and/or disconnection problems associated with the use of storage batteries of 220 V and 6·3 V in circumstances of repeated lightning strikes both to the telephone route and parallel power routes (330 kV Kariba lines), interest was turned to alternating-current feeding. The use of 50 c/s proved impracticable because of too high a longitudinal induction into the telephone pairs on the route which affected direct-current dialling on the many lines so operated. Frequency was then reduced to 15·85 c/s by means of motor alternator sets and transmitted longitudinally to the power feeding pair. The induced voltage at this frequency into other pairs has not proved troublesome. Furthermore, the harmonic content is such that, even with typical 40 dB rejection of longitudinal induction to transverse interference, no difficulties have been experienced. The low value of harmonic content has been achieved by feeding a constant current to line, so obviating any interruption due to frequent operation of the lightning protection devices (silicon carbide) during the rainy season. The circuit is of the Boucherot type which, on a short circuit appearing on the output, limits the current to that determined by inductance. On an open-circuit condition of the power feeding pair, e.g. the wire breaking and falling to the ground, two conditions take effect:

(*a*) A tone generated by the feed current at the last repeater ceases to return and thereby removes the applied power after a short delay.

(*b*) The gap discharger actuates at the sending end of the circuit to restrict the voltage rise, due to the Boucherot circuit, before operation as described in (*a*).

At repeater stations an inverse Boucherot circuit is used to provide constant voltage for rectification and charging of 220 V and 6·3 V lead-acid batteries. Sufficient margin is available for several-days' operation without power feeding. The batteries are contained in a ventilated cupboard in the repeater hut.

6.3. *Supervisory Facilities*

Supervisory facilities include the generation of two tones at each repeater station. The frequencies lie in the 312–328 kc/s range, i.e. the lower four channels in the basic super-group. The tones are injected, one at the input of each amplifier in each direction and one at the output of each amplifier and at appropriate levels. At both Salisbury and Kitwe in the receiving direction 96 voice-frequency telegraph filters are arranged in the four channels lying between 312–328 kc/s after demodulation. The resulting measurements

of tone levels give a location of the majority of faults likely to occur.

6.4. *Gain Adjusting, Alarm and Repeater Switching Facilities*

A single pilot frequency at 308 kc/s is generated at Salisbury, transmitted to Kitwe and returned after amplification, stabilization and filtration. Suitably calculated gain-adjustment networks give appropriate adjustment in 'slope' as well as gain at the repeater stations (shown in Fig. 14).

Normal valve alarm facilities are available and a valve failure removes the supervisory tone from the output of the amplifier concerned, thus giving an indication at the receiving terminal if the system is still working; the use of paralleled valves allows a continuation of operation. The frequency injected at the amplifier input is detected at the output by 'zero-beat' methods and holds a relay in operation. Suitable contacts on the relay disconnect the repeater from the circuit in the event of a failure in the repeater. The gain adjusting repeaters then make a suitable adjustment under the control of the 308 kc/s pilot.

7. Use of Extended Frequency Range

As mentioned in connection with provision of short-haul additional circuits between Kitwe and Ndola, the position in Southern Rhodesia is more favourable due to the standard 40-poles-per-mile spacing throughout.

Naturally a great deal more interference in the broadcast band lying between 564 and 1052 kc/s is to be expected, particularly at night. Measurements taken near Salisbury where there are two broadcast stations setting up field strengths of between 20 mV/m at 5 miles on 584 kc/s to 5 mV/m at 9 miles on 890 kc/s, show interference at −50 dBm at 890 kc/s. In this higher frequency range near-end crosstalk attenuation is better than 50 dB. The indications are, therefore, that for the eighty miles to the first large town, Sinoia, it will be possible to provide an additional 120 channels with modifications to the present equalizers and gain adjusting networks only. Better than 28 dB 'echo' attenuation and a minimum of 50 dB signal/noise ratio, with the use of compandors on the two or three channels receiving significant broadcast interference, will be attained.

8. Cost of System

Figure 15 shows in diagram form the total cost of the system including all channelling equipment to the extent depicted on the channel provision schedule reproduced as Fig. 16.

It is interesting to compare the cost of providing a group of 120 channels over a distance of 305 miles (Salisbury–Lusaka) by means of an 'h.f.' route such as that described and a system using conventional LJ/40-route plus carriers or even a 7400 Mc/s microwave system.[6] For the purposes of the comparison the additional cost of providing a separate pole route instead of using an existing route to carry the 'h.f.' pairs has been taken into account.

120 channels by 'h.f.' route for 305 miles costs £2305 per channel.

120 channels by LJ/40 route for 305 miles costs £4680 per channel.

120 channels by 7400 Mc/s microwave for 305 miles costs £3500 per channel

9. Conclusions

The system is practical and economical for small administrations and despite fears to the contrary the maintenance of a four-wire circuit has not proved particularly difficult during the initial period of 18 months. This is due probably in no small measure to the use of copper-covered steel wire in ideal climatic conditions.

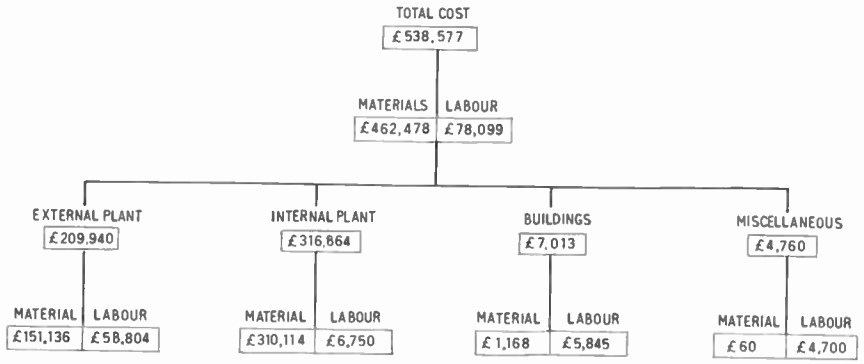


Fig. 15. Cost schedule for the Salisbury–Kitwe 550-mile (880 km) 4-wire h.f. route.

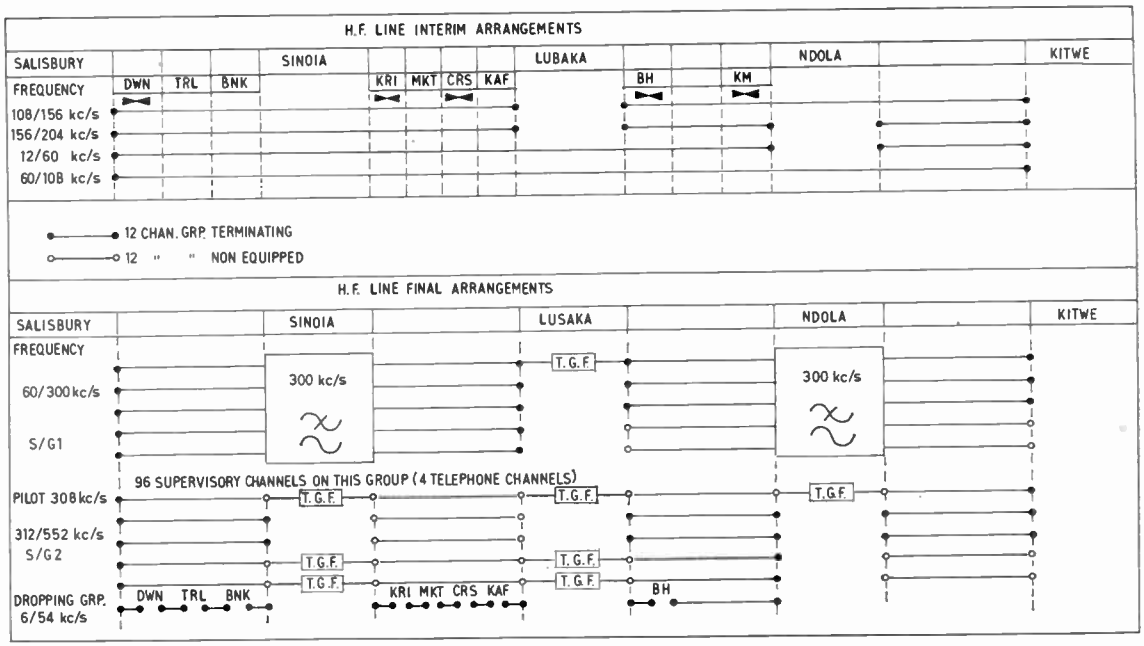


Fig. 16. Channel provision schedule.

10. Acknowledgments

Our thanks are due to our colleagues who co-operated in the many aspects of the design and construction of this system. In particular we wish to thank both Mr. W. F. Cattle, present Chief Engineer, Ministry of Posts, Southern Rhodesia, and Mr. A. Waldram, Director of Telecommunications, Northern Rhodesia, for permission to publish this paper and also for their help and advice throughout the project.

The design and supply of all the line amplifier and associated equipment was carried out by Standard Telephones and Cables Ltd.

11. References

1. M. Vos and C. G. Aurell. "Methods for increasing crosstalk attenuation between overhead lines," *Ericsson Technics*, pp. 113–149, 1936.
2. C.C.I.T.T., Red Book, Vol. III, Recommendation G.122, page 19.
3. E. Baguley and F. B. Cope, "A pulse-echo test set for the quality control and maintenance of impedance uniformity of coaxial cables," *P.O. Elect. Engrs J.*, **44**, Part 4, pp. 164–8, January 1952.
4. F. F. Roberts, "A pulse set for the measurement of small impedance irregularities in high-frequency cables," *J. Instn Elect. Engrs*, **96**, Part III, pp. 17–23, 1949.
5. L. M. Ilgenfritz, R. M. Hunter and A. L. Whitman, "Line problems in the development of the twelve-channel open-wire carrier system," *Bell Syst. Tech. J.*, **18**, pp. 363–37, April 1939.
6. C. F. Boyce, "Open-wire Carrier Telephone Transmission" (McDonald Evans, London, 1962).

Manuscript received by the Institution on 6th January 1964. (Paper No. 919.)

U.D.C. 621.372.852.22

A Perturbation Formula for a Ferrite-loaded Helix

By

K. P. IVANOV†

Summary: The non-reciprocal properties of a thin-wire helix with close windings surrounded by a thin-walled ferrite tube magnetized circumferentially to saturation are examined by means of first-order perturbation theory.

1. Introduction

The problem of electromagnetic wave propagation in a helix surrounded by an unbounded gyromagnetic medium magnetized circumferentially was treated first by Suhl and Walker.[1] The characteristic equation involving higher transcendental functions presents some difficulties in analysing the wave propagation. Several attempts[2, 3] to exploit the non-reciprocal ferrite-loaded helix as an isolator reducing the reflections from the output end in a travelling-wave tube have been reported previously. It is interesting to examine the non-reciprocal properties of the ferrite-loaded helix by means of first order perturbation theory. In the present paper a perturbation formula is derived for a thin-wire helix with close windings surrounded by a thin-walled gyromagnetic tube magnetized circumferentially. Apart from the limited accuracy of the perturbation method the solution obtained provides a satisfactory qualitative idea of the phenomena.

2. Derivation of the Formula

The unperturbed structure is a tightly-wound thin-wire helix with radius a, pitch s (pitch angle $\psi = \arctan s/2\pi a$) propagating a monochromatic electromagnetic wave in the positive z-direction of the appropriate cylindrical co-ordinate system (r, θ, z) (Fig. 1). Assuming a z-dependence of the form $\exp(-j\beta_0 z)$ and taking the sheath model of the helix, the field components of an electromagnetic wave with axial symmetry $(\partial/\partial\theta = 0)$ are expressed in terms of modified Bessel functions[4]

$$E_{r0} = jA(\beta_0/\zeta)I_1(\zeta r) \qquad H_{r0} = jC(\beta_0/\zeta)I_1(\zeta r)$$
$$E_{\theta 0} = -jC(\kappa Z_w/\zeta)I_1(\zeta r) \qquad H_{\theta 0} = jA(\kappa/\zeta Z_w)I_1(\zeta r)$$
$$E_{z0} = AI_0(\zeta r) \qquad H_{z0} = CI_0(\zeta r) \qquad \text{.....(1)}$$

in the region inside, and

$$E_{r0} = -jB(\beta_0/\zeta)\kappa_0(\zeta r) \qquad H_{r0} = -jD(\beta_0/\zeta)\kappa_1(\zeta r)$$
$$E_{\theta 0} = jD(\kappa Z_w/\zeta)\kappa_1(\zeta r) \qquad H_{\theta 0} = -jB(\kappa/\zeta Z_w)\kappa_1(\zeta r)$$
$$E_{z0} = B\kappa_0(\zeta r) \qquad H_{z0} = D\kappa_0(\zeta r) \qquad \text{.....(2)}$$

in the region outside the helix. A, B, C, D are arbitrary constants, $\beta_0 = (\zeta^2 + \kappa^2)^{\frac{1}{2}}$, β_0 and κ are phase constants of the helix and the free space respectively‡ and $Z_w = 120\pi$ is the wave impedance of free space.

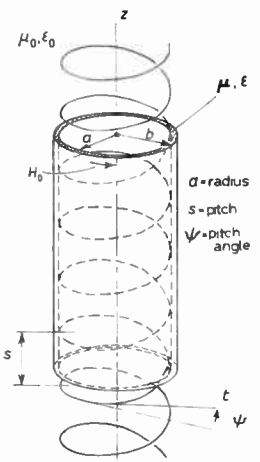


Fig. 1. Tightly-wound helix in a thin-walled ferrite tube.

Applying the boundary conditions of the helically conducting sheet, the coefficients B, C and D are easily expressed in terms of A,

$$B = A\frac{I_0(\zeta a)}{\kappa_0(\zeta a)} \qquad C = -jA\frac{\zeta}{\kappa Z_w}\cdot\frac{I_0(\zeta a)}{I_1(\zeta a)}\tan\psi$$
$$D = jA\frac{\zeta}{\kappa Z_w}\cdot\frac{I_0(\zeta a)}{\kappa_1(\zeta a)}\tan\psi$$

The phase constant β_0 is found from the characteristic equation of the unperturbed helix[4]

$$(\zeta a)^2 = (\kappa a)^2\cot^2\psi\frac{I_1(\zeta a)\kappa_1(\zeta a)}{I_0(\zeta a)\kappa_0(\zeta a)} \qquad \text{.....(3)}$$

The perturbation consists in introducing a thin-walled ferrite cylinder with inner radius a and outer radius b magnetized circumferentially to saturation. The permittivity of the ferrite is a scalar quantity $\varepsilon_0\varepsilon_r$ and the magnetic permeability tensor is of the form (using m.k.s.A. system).

$$\mu = \mu_0\begin{bmatrix}\mu & 0 & -j\alpha\\ 0 & 1 & 0\\ j\alpha & 0 & \mu\end{bmatrix}$$

The shift in phase constant on perturbing the cross section of the helix is easily calculated by means of first order perturbation theory[5, 6]:

† Institute of Electronics, Bulgarian Academy of Science, Sofia, Bulgaria.

‡ The term ζ (in the argument of the modified Bessel functions) is defined by means of β_0 and κ only.

$$\beta-\beta_0=\frac{\omega\int_{S_f}(\Delta\mu\bar{H}.\bar{H}_0^*+\Delta\varepsilon\bar{E}\bar{E}_0^*)\,dS}{\int_{S_0}(\bar{E}_\perp\times\bar{H}_{0\perp}^*+\bar{E}_{0\perp}^*\times\bar{H}_\perp)\bar{e}_z\,dS} \quad ...(4)$$

where $\bar{E}_0$, $\bar{H}_0$, $\bar{E}$, $\bar{H}$ are the fields before and after perturbation, $\bar{E}_{0\perp}$, $\bar{H}_{0\perp}$, $\bar{E}_\perp$, $\bar{H}_\perp$ their transverse components; $\Delta\mu=\mu-\mu_0[1]$, $\Delta\varepsilon=[\varepsilon]-\varepsilon_0[1]$, ([1] is the unit diagonal tensor), $S_f=\pi(b^2-a^2)$ the cross-section of the ferrite tube and S_0 the cross-section of the helix; $\bar{e}_z$ is a unit vector in the direction of z-axis.

Using the quasi-static approximation the internal fields in the gyromagnetic tube are found to be

$$\left.\begin{aligned} H_r&=\frac{1}{\mu}H_{r0}+j\frac{\alpha}{\mu}H_{z0} \quad H_\theta=H_{\theta 0} \quad H_z=H_{z0}\\ E_r&=1/\varepsilon_r E_{r0} \quad\quad E_\theta=E_{\theta 0} \quad E_z=E_{z0}\end{aligned}\right\} \quad ...(5)$$

where the unperturbed fields are those outside the helix.

Since S_f is negligible compared with S_0, the contribution from the tube cross-section and its neighbourhood to the whole integral over S_0 is small. Replacing the perturbed fields in the denominator of eqn. (4) by their unperturbed values we get

$$\int_{S_0}(\bar{E}_{0\perp}\times\bar{H}_{0\perp}^*+\bar{E}_{0\perp}^*\times\bar{H}_{0\perp})\bar{e}_z\,dS=4P_0$$

P_0 being the average power flow through the cross-section of the helix[4]

$$P_0=|A|^2\frac{\pi a^3\kappa}{2Z_w}\cdot\frac{1}{(\beta_0 a)^2}\cdot\frac{I_0}{\kappa_0}\left[\frac{I_1}{I_0}-\frac{I_0}{I_1}+\frac{\kappa_0}{\kappa_1}-\frac{\kappa_1}{\kappa_0}+\frac{4}{\beta_0 a}\right] \quad .....(6)$$

The argument ζa of all Bessel functions is omitted for convenience. For a helix with close windings cotan $\psi > 10$ and $\zeta\simeq\beta_0$ with an error $< 0{\cdot}5\%$.

Substituting from eqn. (5) and eqn. (6) in eqn. (4) and evaluating the integral we get the perturbation formula for a ferrite-loaded helix

$$\Delta\beta a=0{\cdot}5(\beta_0 a)^2\left(\frac{b^2}{a^2}-1\right)\frac{\left(\varepsilon_r-\frac{1}{\varepsilon_r}\right)I_0\kappa_0+I_1\kappa_0\left[\frac{\mu-1}{\mu}\cdot\frac{\kappa_1}{\kappa_0}+\frac{2\alpha}{\mu}+\left(\mu-1-\frac{\alpha^2}{\mu}\right)\frac{\kappa_0}{\kappa_1}\right]}{\left[\frac{I_1}{I_0}-\frac{I_0}{I_1}+\frac{\kappa_0}{\kappa_1}-\frac{\kappa_1}{\kappa_0}+\frac{4}{\beta_0 a}\right]} \quad .....(7)$$

If $\beta_0 a > 3$ the asymptotic expressions for Bessel functions give

$$I_0\kappa_0\simeq\frac{1}{2\beta_0 a},\qquad I_1\kappa_0\simeq\frac{1}{2\beta_0 a},\qquad \frac{\kappa_1}{\kappa_0}\to 1,\qquad \frac{\kappa_0}{\kappa_1}\to 1$$

and the expression in the brackets in the denominator approaches $2/\beta_0 a$. Thus eqn. (7) takes the form

$$\Delta\beta a=0{\cdot}125(\beta_0 a)^2\left(\frac{b^2}{a^2}-1\right)\left[\varepsilon_r-\frac{1}{\varepsilon_r}+\frac{\mu^2-(\alpha-1)^2}{\mu}\right] \quad .....(8)$$

In the case of a polarizing field well below the ferromagnetic resonance ($\mu\simeq 1$, $|\alpha|=\gamma M_s/\omega$) the perturbation formula of a ferrite-loaded helix becomes

$$\Delta\beta a=0{\cdot}125(\beta_0 a)^2\left(\frac{b^2}{a^2}-1\right)\left[\varepsilon_r-\frac{1}{\varepsilon_r}+\alpha(2-\alpha)\right]$$

Inserting $\alpha=0$ in eqn. (8) we get the change in phase constant of a helix perturbed with a thin-walled dielectric tube of relative permittivity ε_r

$$(\Delta\beta a)_d=0{\cdot}125(\beta_0 a)^2\left(\frac{b^2}{a^2}-1\right)\left(\varepsilon_r-\frac{1}{\varepsilon_r}\right)$$

3. Conclusion

From eqn. (7) it is evident that introducing a circumferentially magnetized ferrite tube outside the helix causes a shift in phase constant of the slow electromagnetic wave. Since the off-diagonal element of the permeability tensor depends on the external polarizing field a reversal of this field changes the sign of α. This gives rise to a non-reciprocal effect, i.e. the quantity $\Delta\beta$ becomes different for forward and backward direction of propagation.

If $|\alpha| > |\mu|$, as it may be over a range of values of the external polarizing field, the shift in phase constant of the ferrite-loaded helix becomes less than in the case of dielectric loaded. This means that the wavelength in a helix with ferrite loading is greater than in dielectrically loaded case.

4. References

1. H. Suhl and L. R. Walker, "Topics in guided wave propagation through gyromagnetic media", *Bell Syst. Tech. J.*, 33, pp. 939–4, 1954.
2. J. S. Cook, R. Kompfner and H. Suhl, "Nonreciprocal loss in travelling-wave tubes using ferrite attenuators", *Proc. Inst. Radio Engrs*, 42, pp. 1188–9, 1954.
3. B. N. Enander, "A new ferrite isolator", *Proc. Inst. Radio Engrs*, 44, pp. 1421–30, 1956.
4. W. J. Kleen, "Electronics of Microwave Tubes", Chap. 18, (Academic Press, New York and London, 1958).
5. B. Lax, "Frequency and loss characteristics of microwave ferrite devices", *Proc. Inst. Radio Engrs*, 44, pp. 1368–86, 1956.
6. V. V. Nikol'ski, "Gyromagnetic perturbation of a waveguide", *Radiotekhnika i Elektronika*, 2, pp. 157–71, 1957.

Manuscript first received by the Institution on 16th March 1964 and in final form on 15th May 1964. (Contribution No. 80.)

U.D.C. 621.396.962.3/963.3

The Combination of Pulse Compression with Frequency Scanning for Three-dimensional Radars

By

K. MILNE, Ph.D., B.Sc.(Eng.)†

Reprinted from the Proceedings of the Symposium on "Signal Processing in Radar and Sonar Directional Systems", held in Birmingham on 6th–9th July, 1964.

Summary: Within-pulse frequency scanning can be combined with pulse compression to achieve long-range radar performance combined with high values of data rate, range resolution and height resolution.

The basic principles of a three-dimensional radar of this type are discussed. Scanning is achieved by feeding a beam-squinting array from a frequency-modulated transmitter and pulse compression is applied to all received signals. Height resolution depends upon the aerial's beamwidth and range resolution depends upon the aerial group delay. The concept of a dispersive aerial is introduced and it is shown that this can be used to reduce the bandwidth requirements.

Possible design procedures are examined and an example is given of an S-band radar design which has a high data rate and is virtually free from precipitation clutter effects. Future applications are discussed briefly.

Mathematical appendices analyse the more important spectral relationships of this type of radar.

1. Introduction

The techniques of pulse compression[1] and of frequency-scanning aerials[2] are well-established, and the purpose of the present paper is to show how these two techniques may be combined conveniently for the design of a novel type of three-dimensional pulsed radar. The proposed radar employs 'within-pulse' frequency-scanning in the vertical plane, so that all elevations are examined during each pulse. Elevation information is derived from the 'carrier' frequency of the returned signal, and pulse compression is applied to all received signals to increase the range resolution, thereby improving the performance against clutter.

The present concept arose from considerations of the basic requirements of maintaining a high data rate and of detecting small targets at long ranges and in precipitation clutter. A high data rate implies that all elevation positions be examined during every pulse. The conventional solution employing a single shaped transmitter beam with multiple receiver beams, or the solution employing an electronically scanned receiver beam,[3, 4] provides discrimination against elevation side-lobes on reception only. Frequency scanning provides two-way side-lobe discrimination.

The transmitter energy required to detect a 1 square metre target at 200 miles range with a 1 deg by 1 deg beam is of the order of 100 joules. The maximum value of rain echo obtained with this beamwidth at S-band in precipitation of 4 millimetres per hour is about 10 square metres for every microsecond of pulse length. Thus, even allowing for rain cancellation obtained by the use of circular polarization,[5] the pulse length must be only a fraction of a microsecond if the target is not to be obscured. The necessary peak power for a single-frequency radar is thus several hundreds of megawatts, which is outside the range of current transmitters. A logical solution is to use pulse compression techniques which enable a long transmitter pulse of modest peak power to be compressed on reception to a very short pulse.

The basic principles of a general type of radar are discussed below in Section 2, where it is shown that the radar's coverage can be shaped by controlling the transmitter's spectrum and the aerial group delay. Section 3 discusses possible methods of controlling the transmitter's spectrum and concludes that the most satisfactory method in the present state of the art is to use shaped frequency modulation during a constant amplitude pulse. Section 4 discusses possible design procedures with particular reference to an S-band radar suitable for air-traffic control purposes. The theoretical minimum-power-with-minimum bandwidth design calls for rather impractical aerial and equalizer delay characteristics, but only modest increases in power and bandwidth are required to bring the design within the range of current techniques. Some of the more important practical aspects

† Decca Radar Ltd., Cowes, Isle of Wight.

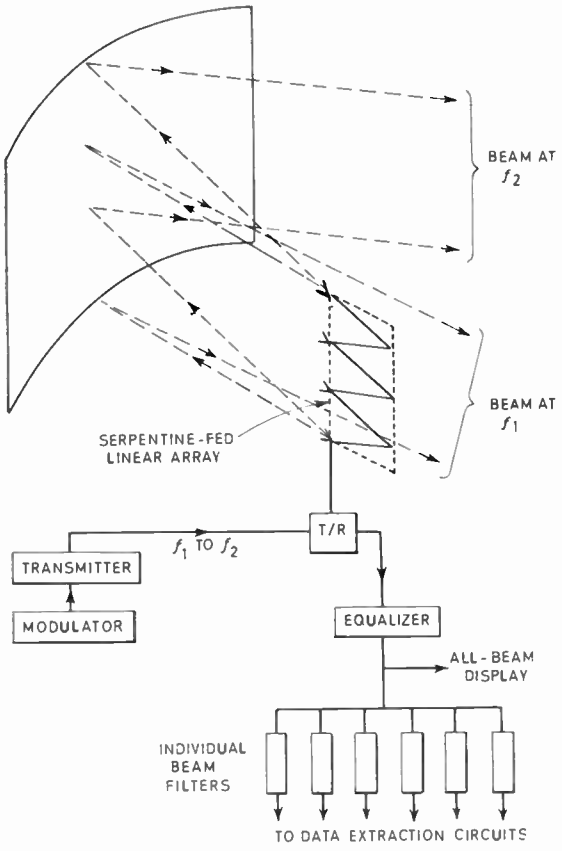


Fig. 1. Schematic diagram of radar.

of the radar are described in Section 5 and finally Section 6 points the way to more sophisticated applications.

2. Basic Principles

The essential components of a general radar of the proposed type are illustrated in Fig. 1. The transmitter may be modulated in both amplitude and frequency. The aerial comprises a conventional cylindrical parabola for forming the horizontal beam and a linear array with serpentine feed[2, 6] for the vertical beam. The elevation angle thus depends upon frequency and by appropriately shaping the transmitter spectrum the power density can be controlled at all elevations to obtain the desired coverage pattern. In the general case, dispersion is introduced into the serpentine feed to provide a non-linear angle-versus-frequency characteristic: this gives an additional degree of freedom to the design which may be used to reduce the total bandwidth required.

Received signals are passed through an equalizer which is an all-pass filter network which compensates for the delay of the transmitter spectrum. (Transmitter delay is easy to visualize when the transmitter frequency is changing with time. A more rigorous definition of delay, based upon the transmitter spectrum, is given in Section 12.) Thus all frequencies tend to arrive simultaneously at the output of the equalizer. Now the principal frequency components which are received from a given elevation angle are those lying within one beamwidth of the aerial and it is shown in Section 11 that these components occupy a bandwidth of $1/D$ where D is the one-way group delay along the serpentine feed.

If the amplitude of the transmitter spectrum is reasonably constant over a frequency interval corresponding to three or four beamwidths, the shape of the received spectrum is determined principally by the aerial: specifically it is the square of the aerial's spectrum at the angle of interest. As shown in Appendix 4, the shape of the output pulse is then given by the convolution of the aerial's aperture illumination. For example, a uniformly illuminated aerial gives rise to a triangular-shaped pulse. The pulse duration is twice the aerial delay. In the ideal

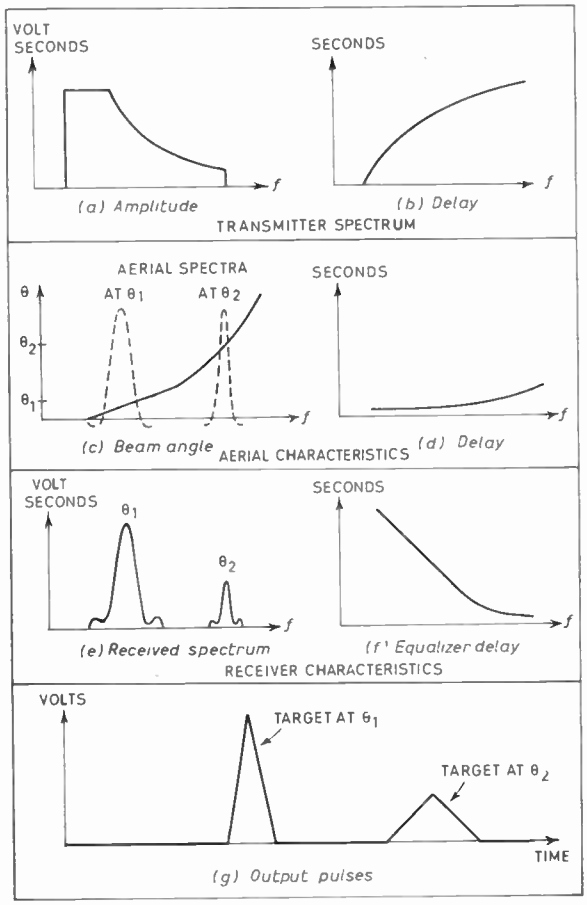


Fig. 2. Basic principles.

case, the output fed to the 'all-beam' display is free from the time side-lobes which appear in conventional pulse-compression radars: the aerial acts as an ideal weighting filter. In practice, ripples appear in the transmitter spectrum principally due to the transients caused by switching on and off (see Appendix 3), and these give rise to small time side-lobes.

Elevation information is obtained by filtering off the appropriate part of the receiver's frequency band. In this case, the final spectrum is affected by the filter response and the exact location of the signal spectrum within the filter's pass-band.

Figure 2 summarizes the basic principles. Figure 2(*a*) shows the amplitude of the transmitted spectrum suitably shaped to reduce the power radiated at frequencies corresponding to the higher elevations. Figure 2(*b*) shows the 'delay' spectrum which is proportional to the slope of the phase of the transmitter spectrum. Figure 2(*c*) illustrates the variation of beam angle with frequency, suggesting some increase in slope at the higher frequencies, implying modest dispersion in the serpentine. The aerial spectrum is also shown at two example angles: the spectral width is reduced at the higher frequency where the aerial delay is greater. The slope of curve (*c*) yields the aerial delay characteristic of Fig. 2(*d*). The spectrum of the signal observed at the target is obtained by multiplying the amplitude of the transmitter spectrum (*a*) by the aerial spectrum (*c*): the delay spectrum is the sum of (*b*) and (*d*). The received spectrum shown in Fig. 2(*e*) involves a further multiplication by the aerial spectrum. Finally, the delay equalizer (*f*) matches the transmitter delay (*b*) so that the output spectrum is essentially that shown at (*e*) with a constant delay. Figure 2(*g*) shows the shape of the output pulse corresponding to (*e*), and this is proportional to the convolution of the aerial's aperture illumination.

The essential feature of the radar is that frequency is synonymous with elevation angle. Angular resolution depends upon the aerial beamwidth and range resolution depends upon the aerial delay D. The transmitter spectrum is shaped to give the required coverage and a bandwidth of $1/D$ is required for each beamwidth of elevation coverage.

Considerations of aircraft target length set a minimum value of D at around 0·1 μs. Smaller values need wider transmission bandwidths but do not usually yield finer range resolution due to the finite size of the target. Considerations of precipitation clutter returns set a maximum value of D at perhaps 0·1 μs again for long ranges and low elevations, rising to several microseconds at higher elevations.

Since frequency is a measure of elevation, a variety of interesting receiver techniques can be devised for the radar. For example, a variable bandwidth filter controlled from range data may be used to reduce the receiver sensitivity at short ranges and low elevations so as to suppress permanent echoes. Similarly, a variable bandwidth, variable centre-frequency filter may be used to abstract height layer information from the receiver.

3. Controlling the Transmitter Spectrum

The modulating function for the transmitter and the resulting spectrum are directly analogous to the aperture illumination and radiation pattern of an aerial. Possible methods of controlling the spectrum, deduced from classical aerial theory[7, 8] are described below.

3.1. *Direct Frequency and Amplitude Modulation of an Oscillator*

The necessary modulating waveforms can be derived from the Fourier Transform of the desired spectrum. In theory, the phase of the transmitter spectrum can be completely arbitrary (provided the equalizer can be designed to suit it!) and hence there are an infinite number of possible solutions. Unfortunately, all solutions require the modulating functions to exist for infinite time. Practical approximations for the modulating waveforms may be obtained by evaluating the instantaneous frequency and envelope of the complex waveform obtained by the method described below.

3.2. *Generation of a Number of Spectral Lines by Separate Oscillators*

This is the analogue of Woodward's method of aperture synthesis.[9] The oscillators are assumed to be pulsed for a time T and the frequency spacing is chosen to be $1/T$. The spectrum from each oscillator is of $(\sin x)/x$ shape with a bandwidth of $1/T$ (see Appendix 3). The complete spectrum can thus be accurately controlled at the carrier frequency of each oscillator by adjusting the individual amplitudes. Elsewhere, ripples of perhaps ± 15% will appear. The various oscillators require to be 'phase-locked' and the phases can be adjusted to give substantially constant delay for all lines, thus dispensing with the need for an equalizing network.

For shaped beams, this method generally results in very peaky amplitude distributions for the combined voltage waveform. Essentially, it tends to produce a pulse of duration very much shorter than T, with low amplitude 'tails' extending over the interval T. Consequently, very low efficiencies will result if a single power amplifier stage is used in the transmitter.

3.3. *Frequency Modulation during a Constant Amplitude Pulse*

This is a special case of the method described in Section 3.1, and appears to be the only practicable

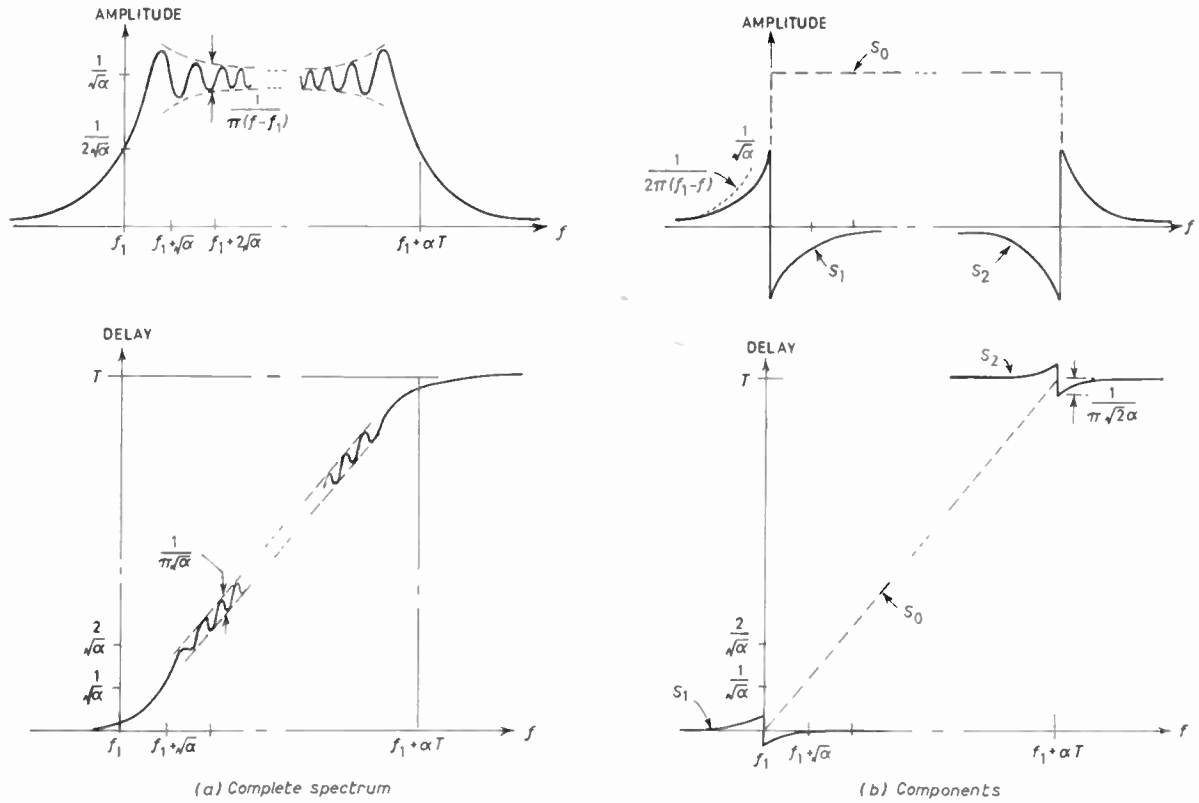


Fig. 3. Spectrum due to linear frequency modulation.

method of obtaining high efficiencies with current transmitter techniques. It is a direct analogy of the geometrical optics approach to the design of shaped-beam aerials using a single feed and a specially shaped reflector.[7]

The power radiated over a given angular interval is assumed to be proportional to the time which the instantaneous transmitter frequency spends in the corresponding frequency interval. The amplitude of the spectrum is thus inversely proportional to the square root of the rate of change of transmitter frequency, and the necessary frequency modulating waveform is obtained directly from the desired coverage diagram.

The actual spectrum is the Fourier transform of the transmitted waveform, and will inevitably contain minor fluctuations. Details of a method of computing the spectrum for an arbitrary frequency modulation are given in Appendix 3. The simple case of linear f.m. is analysed in some detail, and is illustrated by Fig. 3. The fluctuations in the spectrum can be identified with transient effects due to switching the transmitter on and off. This is made evident in Fig. 3, where the spectrum is separated into three components. The main component is of constant amplitude and has the linear delay characteristic of the instantaneous transmitter frequency. The components due to switching are centred around the frequencies corresponding to the edge frequencies of the transmitter, and have substantially constant delay spectra.

The fluctuations in the transmitted spectrum give rise to time side-lobes in the receiver output. In the

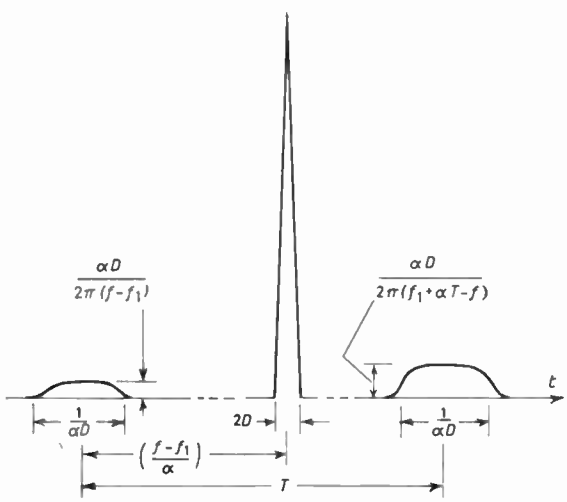


Fig. 4. Amplitudes of spurious output pulses (linear f.m.).

particular case of linear f.m., it will be evident that the principal effect is to produce subsidiary responses at times corresponding to the edges of the transmitter pulse. The effect is analysed in Appendix 4 (Sect. 13.4), and is illustrated in Fig. 4. The amplitudes of the subsidiary responses decrease if the transmitter bandwidth is increased.

In the case of non-linear f.m., further fluctuations in the spectra can be anticipated, particularly if there are discontinuities in the slope of the transmitter's frequency deviation versus time characteristic. By analogy with shaped-beam aerials, we may expect fluctuations of about ± 20% in amplitude and ± 15 deg in phase. A rough estimate suggests that such fluctuations will produce a time side-lobe level not exceeding − 20 dB. Whilst it is theoretically possible to reduce the effects of the fluctuations by introducing compensating amplitude and phase shaping into the equalizer, such a procedure is likely to call for extremely accurate control of both transmitter and equalizer and is probably beyond the present state of the art.

4. Design Procedures for Constant Amplitude F.M. Systems

4.1. *General*

The basic design procedure consists of treating each elevation beamwidth as a single-frequency radar problem. The necessary steps are as follows:

(*a*) Determine the minimum permissible transmitter pulse length for each beam position from the required coverage diagram and the minimum target size. Integration of the pulse length per beam yields the total pulse length and hence the mean transmitter power.

(*b*) Determine the range resolution required as a function of elevation. This will often be dictated by consideration of rain clutter and some reduction in resolution may be permissible at the higher elevations. Range resolution determines the maximum permissible compressed pulse length and hence the aerial delay. Integration of the reciprocal of aerial delay yields the total bandwidth required.

As a typical example of the design approach, consider an S-band air-traffic control radar which is required to give 80% detection probability for a 2 metre^2 target at ranges up to 200 nautical miles and in precipitation with a ceiling of 20 000 feet and rates of up to 4 mm/hr. The following parameters are assumed:

Peak transmitter power, P_T	2·5 MW
Frequency, f	3 Gc/s
Aerial gain (per beam), G	47·5 dB
Horizontal beamwidth, Φ	$\frac{1}{2}^\circ$
Vertical beamwidth, Θ	1°
Vertical coverage	0° to 20°
Maximum range, R_{max}	200 nautical miles
Minimum target size, σ	2 m^2
Pulse repetition frequency, p	250 c/s
Data rate	10 rev/min
Pulses per horizontal beamwidth, N	2·1
Polarization	Circular
Rain cancellation ratio[5], K	18 dB
Receiver noise temperature, T_0	1500°K
System losses, L	
scanning loss, L_s	1·8 dB
polarization loss, L_p	2·5 dB
waveguide and filter loss, L_f	2·2 dB
rain and atmospheric attenuation (at maximum range), L_a	3·5 dB
	10 dB

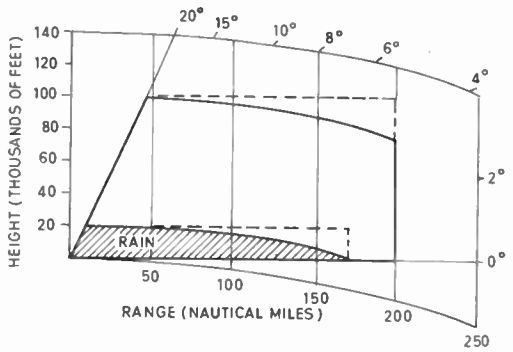


Fig. 5. Desired coverage.

The required coverage diagram is illustrated in Fig. 5. The shaded area represents the extent of the coverage affected by rain clutter. For simplicity, the coverage diagram shown by the dotted line is assumed for the remainder of the paper.

4.1.1. Transmitter pulse length

The minimum detectable signal, S_{min}, may be obtained from curves given by Skolnik.[10] In the present case S_{min} is about 16 dB for 80% probability of detecting a target which is fading slowly with a Rayleigh probability distribution. The minimum transmitter pulse length η may then be found from the usual radar equation,[10] by assuming an effective receiver bandwidth of $1/\eta$. This yields a value of 2 μs for the pulse length per beamwidth at low elevations.

The variation of minimum pulse length per beamwidth with elevation is illustrated in Fig. 6. At the

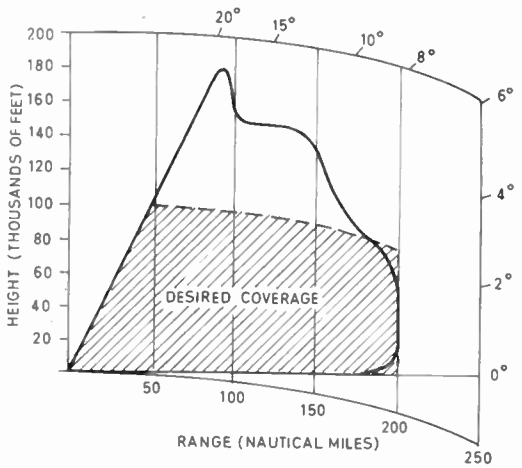
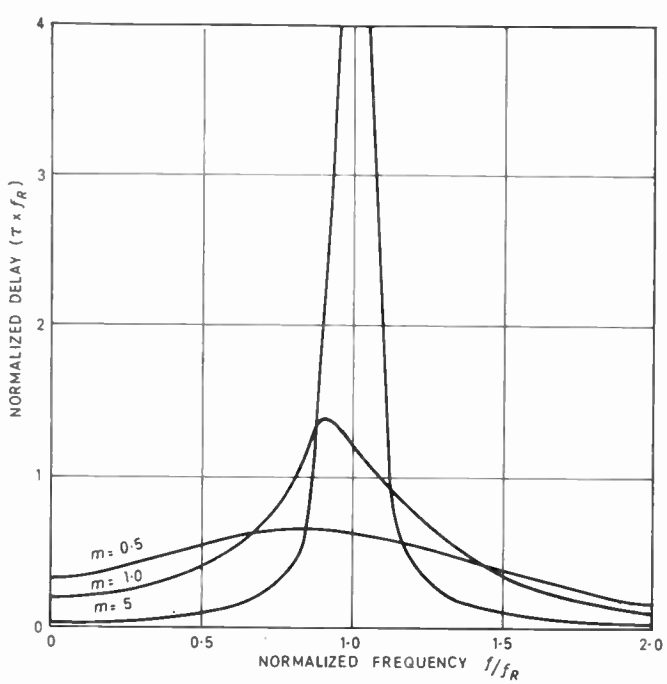


Fig. 9. Coverage of 'practical' design.

described by Allnatt.[12] The delay characteristics of these elements are illustrated in Fig. 10. The parameters m and f_R were chosen as 1 and 166 Mc/s respectively: this yields the required slope of 0·2 μs per Mc/s, and reasonable linearity over the first 50 Mc/s above f_R, with about 2000 sections.

More economical equalizer designs may be obtained by cascading elements having differing values of m and f_R. In this way, the equalizer characteristic may be more closely matched to the desired coverage diagram, but care is needed to avoid producing ripples in the characteristic.



5.2. *Dispersive Aerials*

Dispersion may be introduced into the serpentine feed by distributed reactive loading, or by lumped elements. The microwave equivalent of the lattice element is shown in Fig. 11.

There is not so much freedom of choice in the design of lumped elements for the aerial as in the design of the equalizer. All the elements must be identical since they are situated between the individual radiators. Furthermore, the centre frequency of the element is determined by the radar. For the design described above, f_R is about 3 Gc/s, and a value of $m = 5$ was chosen to give a reasonably linear delay change per element of about 1 nanosecond over an interval of 100 Mc/s. The aerial requires about 100 radiators

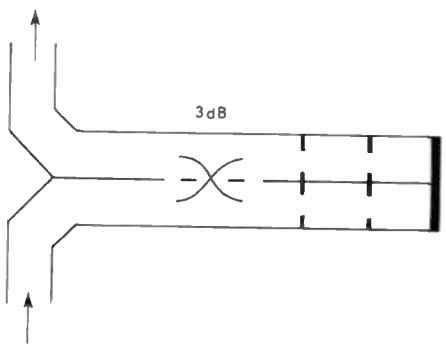


Fig. 11. Microwave dispersive element.

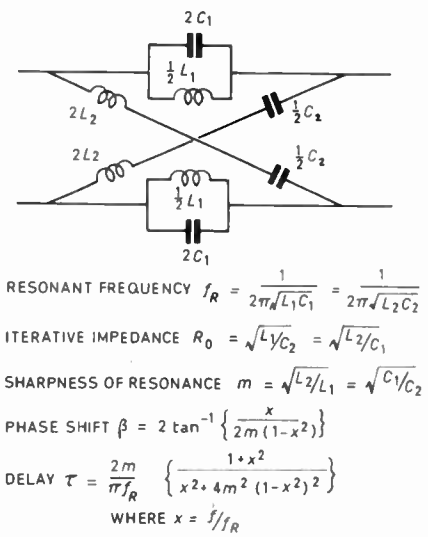


Fig. 10. Delay equalizer element.

spaced at about 0·65 wavelengths to produce a 1-deg beam. Thus one element per radiator gives a delay change of 0·1 μs over a 100 Mc/s bandwidth.

Further work needs to be done to find a method of introducing dispersion with minimal loss.

5.3. *Loss in the All-beam Display*

The previous calculations of range performance implied detection in the individual filters corresponding to each vertical beamwidth. Detection probabilities in the all-beam display will be lower.

At first sight, it may seem that a loss of 13 dB would be incurred due to the addition of noise from 20 beam 'positions'. However, the detected signals fed to the display will be filtered so as to pass only signals of the correct pulse length (0·1 to 0·2 μs in the above example). The 'loss' is thus the loss of post-detector as compared to pre-detector integration. This is approximately $M^{\frac{1}{2}}$ where M is the number of beam 'positions'. At long ranges, only signals from the lower five beam positions need be fed to the all-beam display: the loss will thus be about 2·5 dB.

5.4. *Distortion in the Height Filters*

Section 2 demonstrated that the pulse shape was given by the convolution of the aerial's aperture illumination and was free from time side-lobes. This is true if the 'side-lobes' of the aerial spectrum are allowed to contribute to the output, since the aerial is acting as a weighting filter. In the case of the height filters, the spectral side-lobes are suppressed and time side-lobes may re-appear.

Figure 12 illustrates the distortion introduced by a single-tuned circuit. The pulse shape depends upon the exact location of the signal spectrum in the filter pass-band. The distortion can be reduced by employing filters with Gaussian shaped response curves and substantially linear phase-versus-frequency characteristics.

The height finding accuracy of the radar is expected to be similar to that of a conventional multi-beam radar of the same vertical beamwidth since identical techniques of beam comparison may be applied. One aspect which requires further investigation is the effect of target fading upon the received spectrum: it is anticipated that this will not be serious in those designs where the aerial's spectral width is restricted to about 10 Mc/s.

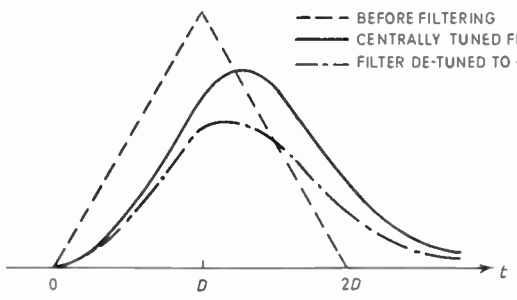


Fig. 12. Pulse distortion due to i.f. filter.

6. Future Possibilities

Whilst the radar design discussed above is intended for air-traffic control purposes, many other possibilities exist for the technique of combining frequency scanning with pulse compression.

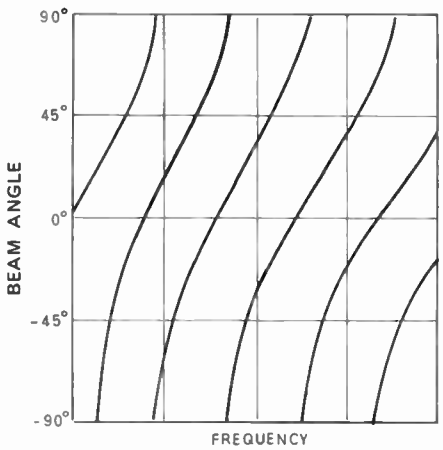


Fig. 13. Typical aerial characteristic.

The extension to inertialess scanning in the horizontal plane by using a stack of serpentine feeds fed via electronically-controlled phase shifters is an obvious one. The limitation here is likely to be an economic one: the current costs of serpentine feeds are high.

An attraction of the frequency scanner is the ability to restrict or even stop the scan to examine a chosen elevation interval in more detail. An additional narrow-bandwidth, large-delay equalizer could be switched in to maintain the same compressed pulse length under these conditions.

In applications where strong interference is encountered, there may be some objection to a fixed angle-versus-frequency relationship. This relationship is in fact periodic, as illustrated in Fig. 13. If wide-band transmitters are available, a number of frequencies may be chosen for the same elevation angle. Croney[6] has suggested feeding alternate radiators from separate serpentines and switching in a phase shift of 180 deg into one serpentine. This doubles the number of possible frequencies. This technique

may be extended by increasing the number of serpentine feeds.

The angle-versus-frequency relationship may also be modified by including electrically controlled phase shifters in each loop of the serpentine feed. The aerial then becomes a combined phase and frequency scanner, relying upon frequency change to provide within-pulse scanning with pulse compression, whilst the phase shifters may be switched on a pulse-to-pulse basis to provide a large number of possible frequencies for each elevation.

7. Conclusions

The principle of combining 'within-pulse' frequency scanning with pulse compression has been established. The technique enables a high data rate three-dimensional radar to be designed which is virtually free from precipitation clutter effects and which offers two-way discrimination against elevation side-lobes. Elevation information is obtained from the frequency of the echo and a variety of receiver techniques can be employed to perform functions such as height layering. All elevations are examined within each pulse and the data rate can be as high as the azimuth beamwidth and pulse repetition frequency will permit. An important advantage of the proposed radar is that a single aerial design may be used for a wide variety of coverage patterns since only the delay equalizer and frequency modulation of the transmitter need be changed.

The concept of a dispersive serpentine aerial has been introduced and it has been shown how this can be used to reduce the bandwidth required for the radar.

Future applications of the technique appear to offer scope for considerable ingenuity in the design of sophisticated radar systems.

8. Acknowledgments

The author is indebted to Mr. J. G. Flounders for a number of stimulating discussions which have led to the present paper.

Acknowledgment is made to the directors of Decca Radar Ltd. for permission to publish the paper.

9. References

1. J. R. Klauder, A. C. Price, S. Darlington and W. J. Albersheim, "The theory and design of chirp radars", *Bell Syst. Tech. J.*, **39**, p. 745, July 1960.
2. H. Shnitkin, "Survey of electronically-scanned antennas", *Microwave J.*, **3**, No. 12, p. 67, December 1960, and **4**, No. 1, p. 57, January 1961.
3. D. E. N. Davies, "A fast electronically-scanned radar system", *J. Brit.I.R.E.*, **21**, p. 301, 1961.
4. D. E. N. Davies, "Application of electronic sector scanning techniques to height-finding radar systems", *Proc. Instn Elect. Engrs*, **110**, No. 11, p. 1941, November 1963.
5. H. Gent, I. M. Hunter and N. P. Robinson, "Polarization of radar echoes including aircraft, precipitation and terrain", *Proc. I.E.E.*, **110**, No. 12, p. 2139, December 1963.
6. J. Croney, "Double-dispersive frequency scanning antenna (for two-plane scanning)", *Microwave J.*, **6**, No. 7, p. 76, July 1963.
7. S. Silver, "Microwave Antenna Theory and Design", Chaps. 6, 9 and 13 (McGraw-Hill, New York, 1949).
8. M. I. Skolnik, "Introduction to Radar Systems", Chap. 7 (McGraw-Hill, New York, 1962).
9. P. M. Woodward, "A method of calculating the field over a plane aperture required to produce a given polar diagram" *J.I.E.E.*, **93**, Pt. IIIA, p. 1554, 1946.
10. Skolnik, *op. cit.*, Chap. 2.
11. W. J. James, "The effect of weather in Eastern England on the performance of X-band ground radars", R.R.E. Technical Note, No. 655, July 1961.
12. J. W. Allnatt, "The delay equalization of the London-Birmingham television cable system", *Proc. I.E.E.*, **99**, Pt. IIIA, No. 18, p. 338, 1952 (Convention on the British Contribution to Television, April–May 1952).
13. T. Pearcey, "Table of the Fresnel Integral to six decimal places" (Cambridge University Press, 1956).
14. H. E. Hawkins and O. La Plant, "Radar performance degradation in fog and rain", *Trans. Inst. Radio Engrs (Aerospace and Navigational Electronics)*, **ANE-6**, p. 26, March 1959.

10. Appendix 1

List of Symbols

a	Length of aerial aperture.
$A(f)$	Amplitude of transmitter spectrum.
B	Bandwidth of i.f. filter.
c	Velocity of electromagnetic waves.
$C(v)$	Fresnel integral.
D, $D(f)$	Total group delay along serpentine aerial (one-way).
$D'(f)$	First derivative of $D(f)$.
$E(\sin\theta, f)$	Radiation pattern of aerial at angle θ and frequency f.
f	Frequency.
f_0	Frequency at which beam is normal to array.
f_1, f_2	Limits of transmitter frequency sweep.
f_m	Mean frequency received at target.
f_c	Centre frequency of i.f. filter.
$F(f)$	Response of i.f. filter.
$G(f)$	Received frequency spectrum after equalization.
$G_1(f)$	Received frequency spectrum before equalization.
$g(x)$	Amplitude distribution at distance x along aerial.
h	Target altitude.

h_1 Rain ceiling.
$h(t)$ Dummy time function used in convolution integrals.
$H(t)$ Receiver output.
$H_1(t)$ Receiver output after i.f. filtering.
K Rain cancellation ratio.
L Loss factor used in computing radar range.
M Number of beam positions.
N Number of pulses per beamwidth.
p Pulse repetition frequency.
P_T Transmitter peak power.
R Range.
$s(f)$ Frequency spectrum.
s_0, s_1, s_2 Components of frequency spectrum.
$S(v)$ Fresnel integral.
T_0 Receiver noise temperature.
T Duration of transmitter pulse.
$u(t)$ Transmitter voltage.
$V(t)$ Response of i.f. filter to Dirac function
v Dummy variable.
x Distance along aerial aperture.
y, z Dummy variables.
α Rate of change of transmitter frequency.
β, γ Second and third derivatives of transmitter frequency.
$\delta(t)$ Dirac function.
η Transmitter pulse length per elevation beamwidth.
ζ Compressed pulse length per elevation beamwidth.
θ Elevation angle.
Θ Elevation beamwidth.
λ Wavelength.
ρ Effective radius of curvature of the the earth.
σ Target echoing area.
σ_R Rain echoing area.
σ_u Specific echoing area of rain (m^2/m^3).
$\tau(f)$ 'Delay' spectrum (defined in Appendix 3).
τ_e Equalizer delay characteritic.
τ_0 Fixed delay in equalizer.
Φ Azimuthal beamwidth.
$\chi(f)$ Phase of transmitter spectrum.
$\psi(f)$ Total phase shift along serpentine aerial.

11. Appendix 2

Radiation Patterns and Frequency Spectra of Linear Arrays

The serpentine array discussed in the text consists of a number of discrete elements, but in discussing radiation patterns it is more convenient to consider the equivalent continuous array. The radiation pattern is then given by the Fourier transform of the aperture illumination.[7, 8] Since the phase distribution along the array is always linear, the radiation pattern at range R and angle θ for an array of length a operating at frequency f is proportional to:

$$E(\sin\theta, f) = \exp\left(-\mathrm{j}\frac{2\pi fR}{c}\right) \times \int_0 g(x)\exp\left\{\mathrm{j}x\left(\frac{2\pi f\sin\theta}{c} + \frac{1}{a}\psi(f)\right)\right\}\mathrm{d}x \quad (1)$$

where $g(x)$ represents the amplitude distribution at distance x (assumed independent of frequency) and $\psi(f)$ represents the total phase shift along the serpentine feed. $\psi(f)$ may be conveniently written as

$$\psi(f) = -2\pi \int_{f_0}^{f} D(v)\,\mathrm{d}v \qquad \ldots\ldots(2)$$

where $D(f)$ is the group delay along the serpentine at frequency f and f_0 is the frequency at which the beam is normal to the array, i.e. the phase distribution along the array is constant; (f_0 should be evaluated for the array of discrete elements, since for a truly continuous array the only possible value of f_0 is zero).

The radiation pattern at frequency f and angle θ may now be written in terms of the pattern at frequency f_0, namely:

$$E(\sin\theta, f) = \exp\left\{-\mathrm{j}\frac{2\pi(f-f_0)R}{c}\right\}.E(\sin\theta_0, f_0) \quad \ldots(3)$$

where

$$E(\sin\theta_0, f_0) = \exp\left\{-\mathrm{j}\frac{2\pi f_0 R}{c}\right\} \times \int_0^a g(x)\exp\left\{\mathrm{j}\frac{2\pi f_0 x\sin\theta_0}{c}\right\}\mathrm{d}x \quad (4)$$

is the pattern at θ_0, f_0 and the angle θ_0 is defined by:

$$\sin\theta_0 = \frac{f}{f_0}\sin\theta - \frac{c}{a.f_0}\int_{f_0}^{f} D(v)\,\mathrm{d}v \qquad \ldots\ldots(5)$$

The pattern at frequency f may thus be obtained from the pattern at frequency f_0 by changing the $\sin\theta$ scale according to equation (5). The location of the peak is clearly given by:

$$\sin\theta_m = \frac{c}{af}\int_{f_0}^{f} D(v)\,dv \qquad \text{.....(6)}$$

Relationships between beam shape and amplitude distribution as determined by equation (4) are given in the literature[7, 8]: the beamwidth varies from $0{\cdot}88\lambda/a$ for uniform illumination to about $1{\cdot}2\lambda/a$ for moderately tapered illuminations.

Differentiation of equation (5) yields an important relationship between rate of change of beam angle with frequency and aerial delay, specifically

$$\frac{d\theta}{df} = \frac{c}{af}\sec\theta\,.\,D(f) - \frac{\tan\theta}{f} \qquad \text{.....(7)}$$

For operation close to normal this reduces to

$$\frac{d\theta}{df} \simeq \frac{\lambda}{a}D(f) \qquad \text{.....(8)}$$

Since the beamwidth is approximately λ/a, the number of beamwidths scanned per cycle per second of frequency change is approximately equal to the aerial delay.

Equation (3) may be interpreted as either the radiation pattern at a fixed frequency or as the frequency spectrum at a fixed angle. In the former case, the location of the radiation peak is given by equation (6) and the beamwidth is related to the beamwidth at f_0 by the factor

$$\frac{\Theta}{\Theta_0} \simeq \frac{f_0}{f}\sec\theta_m \qquad \text{.....(9)}$$

In the latter case, the location of the spectral peak is again given by equation (6). The bandwidth of the spectrum may be related to the beamwidth by the formula

$$\Delta f \simeq \frac{a}{\lambda}\cdot\frac{\cos\theta}{D(f)}\cdot\Theta \qquad \text{.....(10)}$$

or

$$\Delta f \simeq \frac{a}{\lambda_0}\frac{\Theta_0}{D(f)} \qquad \text{.....(11)}$$

and since the beamwidth is approximately λ/a, the bandwidth is approximately equal to the reciprocal of the aerial delay.

12. Appendix 3

The Transmitter Spectrum

The transmitter spectrum $s(f)$ is the Fourier transform of the transmitter voltage $u(t)$, specifically:

$$s(f) = \int_{-\infty}^{\infty} u(t)\exp(-j2\pi ft)\,dt \qquad \text{.....(12)}$$

It is convenient to separate the amplitude and phase components of the spectrum by writing

$$s(f) = A(f).\exp j\chi(f) \qquad \text{.....(13)}$$

where $A(f)$ and $\chi(f)$ are real functions. The concept of 'delay spectrum' may now be introduced by defining

$$\tau(f) = \frac{-1}{2\pi}\frac{d\chi(f)}{df} \qquad \text{.....(14)}$$

The delay spectrum is sometimes easier to envisage than the phase of the spectrum. When a signal is passed through a network, the output delay spectrum is simply the sum of the signal and network delay spectra. The phase of the output spectrum can always be obtained by integrating the delay spectrum (apart from an arbitrary constant).

In the present paper, we are primarily concerned with pulses of constant amplitude and variable frequency. If the instantaneous frequency is represented by a polymonial of the form:

$$f = f_1 + \alpha t + \beta t^2 + \ldots \qquad \text{.....(15)}$$

the expression for the transmitter voltage becomes

$$u(t) = \exp\{j2\pi(f_1 t + \tfrac{1}{2}\alpha t^2 + \tfrac{1}{3}\beta t^3 + \ldots)\} \qquad \text{.....(16)}$$

Evaluation of the spectrum from equation (12) usually involves extensive numerical integration except in the cases of no frequency modulation (α and higher coefficients zero) and linear frequency modulation (β and higher coefficients zero). These cases are analysed below in some detail, since they point the way to a method of computing the spectrum in the more general case.

12.1. *Constant-amplitude, Constant-frequency Pulse*

If the pulse duration is T, then

$$u(t) = \exp\{j2\pi f_1 t\}, \qquad 0 < t < T \qquad \text{.....(17)}$$

and the resulting spectrum is

$$s(f) = T\cdot\frac{\sin\{\pi(f-f_1)T\}}{\pi(f-f_1)T}\cdot\exp\{-j\pi(f-f_1)T\} \qquad \text{...(18)}$$

The amplitude of the spectrum is thus of $\sin x/x$ shape, with a half-power width of approximately $1/T$. The delay spectrum is

$$\tau(f) = -\frac{1}{2\pi}\frac{d\chi(f)}{df} = \frac{T}{2} \qquad \text{.....(19)}$$

which is simply the time delay corresponding to the centre of the pulse.

12.2. *Constant-amplitude Pulse with Linear Frequency Modulation*

In this case,

$$u(t) = \exp\{j2\pi(f_1 t + \tfrac{1}{2}\alpha t^2)\}, \qquad 0 < t < T \quad (20)$$

and the resulting spectrum is

$$s(f) = \frac{1}{\sqrt{2\alpha}}\cdot\exp\left\{-j\frac{\pi(f-f_1)^2}{\alpha}\right\}\times$$
$$\times\{[C(v_1)-C(v_2)]+j[S(v_1)-S(v_2)]\} \quad (21)$$

where

$$v_1 = \sqrt{\frac{2}{\alpha}}\{f-f_1\} \qquad \text{.....(22)}$$

$$v_2 = \sqrt{\frac{2}{\alpha}}\{f-(f_1+\alpha T)\} \qquad \text{.....(23)}$$

and $C(v_1)$ and $S(v_1)$ are the Fresnel integrals defined by

$$C(v_1) = \int_0^{v_1} \cos\left(\frac{\pi}{2}v^2\right) dv \qquad \text{.....(24)}$$

$$S(v_1) = \int_0^{v_1} \sin\left(\frac{\pi}{2}v^2\right) dv \qquad \text{.....(25)}$$

The delay spectrum is

$$\tau(f) = \left(\frac{f-f_1}{\alpha}\right) - \frac{1}{2\pi}\frac{d}{df}\left\{\tan^{-1}\frac{S(v_1)-S(v_2)}{C(v_1)-C(v_2)}\right\} \qquad \text{...(26)}$$

which can be evaluated as:

$$\tau(f) = \left(\frac{f-f_1}{\alpha}\right) - \frac{1}{\pi}\sqrt{\frac{2}{\alpha}}\times$$

$$\times \sin\frac{\pi}{4}(v_1^2-v_2^2)\left\{\frac{\begin{array}{l}[C(v_1)-C(v_2)]\cos\frac{\pi}{4}(v_1^2+v_2^2)+\\ \quad +[S(v_1)-S(v_2)]\sin\frac{\pi}{4}(v_1^2+v_2^2)\end{array}}{[C(v_1)-C(v_2)]^2+[S(v_1)-S(v_2)]^2}\right\}$$

.....(27)

The first term gives the principal contribution in the region $f_1 < f < (f_1 + \alpha T)$. Its value is simply the time taken for the instantaneous transmitter frequency to reach f.

Figure 3 illustrates the amplitude and delay spectra for the case of a large value of the transmitter time-bandwidth product, i.e. $\alpha T^2 \gg 1$. The amplitude of the spectrum in the band of transmitted frequencies is approximately $1/\sqrt{\alpha}$. Outside this band, the amplitude approaches zero and the delay spectrum approaches the values 0 and T.

In order to throw some more light on the nature of the amplitude and delay spectra, it is convenient to re-write equation (21) in the form

$$s(f) = s_0(f)+s_1(f)+s_2(f) \qquad \text{.....(28)}$$

where $s_0(f)$ is the value of the spectrum which would be expected from consideration of the frequency modulation impressed on the transmitter, specifically:

$$s_0(f) = \begin{cases} 0, & f < f_1 \\ \frac{1}{\sqrt{\alpha}}\exp\left\{j\frac{\pi}{4} - j\frac{\pi}{\alpha}(f-f_1)^2\right\}, & f_1 < f < (f_1+\alpha T) \\ 0, & (f_1+aT) < f \end{cases}$$

.....(29)

The functions $s_1(f)$ and $s_2(f)$ may be identified with additions to the expected spectrum caused by switching on and switching off the transmitter if we choose:

$$s_1(f) = \frac{1}{\sqrt{2\alpha}} \cdot \exp\left\{-j\frac{\pi(f-f_1)^2}{\alpha}\right\}\times$$

$$\times\begin{cases}[C(v_1)+\frac{1}{2}]+j[S(v_1)+\frac{1}{2}], & f<f_1\\ [C(v_1)-\frac{1}{2}]+j[S(v_1)-\frac{1}{2}], & f_1<f\end{cases}$$

.....(30)

and

$$s_2(f) = \frac{-1}{\sqrt{2\alpha}} \cdot \exp\left\{-j\frac{\pi(f-f_1)^2}{\alpha}\right\}\times$$

$$\times\begin{cases}[C(v_2)+\frac{1}{2}]+j[S(v_2)+\frac{1}{2}], & f<(f_1+\alpha T)\\ [C(v_2)-\frac{1}{2}]+j[S(v_2)-\frac{1}{2}], & (f_1+\alpha T)<f\end{cases}$$

.....(31)

$s_1(f)$ is thus an odd function centred around f_1 and $s_2(f)$ is a similar function centred around $(f_1+\alpha T)$.

Further insight into the nature of s_1 and s_2 may be obtained by using the asymptotic expansions of the Fresnel integrals[13] namely,

$$C(v)+jS(v) = \tfrac{1}{2}(1+j) -$$

$$-\frac{j}{\pi v}\exp j\frac{\pi v^2}{2}\left\{\begin{array}{l}\left[1-\frac{1.3}{\pi^2}\cdot\frac{1}{v^4}+\ldots\right]\\ \frac{-j}{\pi v^2}\left[1-\frac{1.3.5}{\pi^2}\cdot\frac{1}{v^4}+\ldots\right]\end{array}\right\}$$

.....(32)

Noting that $C(-v) = -C(v)$, the function s_1 can be reduced to

$$s_1(f) = \frac{-j}{2\pi(f-f_1)}\left\{1-\frac{j\alpha}{2\pi(f-f_1)^2}+\ldots\right\} \qquad \text{...(33)}$$

Thus, at a sufficient distance from the edge frequency f_1, this component has zero delay and the amplitude is inversely proportional to $(f-f_1)$. Similarly, the function s_2 can be reduced to

$$s_2(f) = \frac{+j.\exp j\{\pi\alpha T^2 - 2\pi(f-f_1)T\}}{2\pi(f-f_1-\alpha T)}\times$$

$$\times\left\{1-\frac{j\alpha}{2\pi(f-f_1-\alpha T)^2}+\ldots\right\} \qquad \text{...(34)}$$

which has the same shape as s_1 but is centred round $(f_1+\alpha T)$ and has a delay of T.

The components s_0, s_1 and s_2 are illustrated in Fig. 3.

12.3. *Spectrum of Constant-amplitude Pulse with Arbitrary Frequency Modulation*

Some aspects of the spectrum due to an arbitrary frequency modulation may be deduced from the analysis given above for linear f.m. The latter shows that the spectral amplitude is proportional to $1/\sqrt{\alpha}$ where α is the rate of change of transmitter frequency with time. Thus for the general case of the transmitter frequency being given by a polynomial of the form:

$$_1+\alpha t+\beta t^2+\gamma t^3+\ldots \qquad \ldots\ldots(35)$$

we would expect the spectral amplitude to be inversely proportional to the square root of the rate of change of transmitter frequency namely:

$$A(f) \simeq \frac{1}{\sqrt{\alpha+2\beta y+3\gamma y^2+\ldots}} \qquad \ldots\ldots(36)$$

where the parameter y corresponds to the time at which the instantaneous frequency is f, i.e. y is the root of the equation

$$f=f_1+\alpha y+\beta y^2+\gamma y^3+\ldots \qquad \ldots\ldots(37)$$

Furthermore, we would expect the delay spectrum to be given by

$$\tau(f) \simeq y \qquad \ldots\ldots(38)$$

One method of obtaining a better assessment of the spectrum is to divide the pulse duration into a number of intervals over which frequency change is reasonably linear. A suitable interval length would be one in which the phase error introduced by assuming linear f.m. does not exceed $\pm\pi/4$. The contribution to the spectrum from each interval can then be obtained from equation (21). If further accuracy is desired, the phase error may be expanded as a polynomial. The spectrum may then be expressed by equation (21) plus a series involving its derivatives. Specifically, if

$$s(f)=\int u(t)\exp(-\mathrm{j}2\pi ft)\,\mathrm{d}t \qquad \ldots\ldots(39)$$

then

$$\frac{\mathrm{d}^n s(f)}{\mathrm{d}f^n}=\int(-\mathrm{j}2\pi t)^n.u(t).\exp(-\mathrm{j}2\pi ft)\,\mathrm{d}t \ldots(40)$$

Thus if we know the spectrum due to $u(t)$ and we wish to find the spectrum $s_1(f)$ due to

$$u_1(t)=u(t).\exp(\mathrm{j}at^n) \qquad \ldots\ldots(41)$$

we expand the exponential, obtaining

$$u_1(t)=u(t)\left\{1+(\mathrm{j}at^n)+\frac{(\mathrm{j}at^n)^2}{2!}+\ldots\right\} \ldots(42)$$

and the spectrum is

$$s_1(f)=s(f)+\sum_{m=1}^{\infty}\frac{\mathrm{j}^{(n+1)m}a^m}{(2\pi)^{nm}m!}\cdot\frac{\mathrm{d}^{nm}s(f)}{\mathrm{d}f^{nm}} \qquad \ldots(43)$$

If the phase error is small, it will usually be sufficiently accurate to take the first term in the series of derivatives of $s(f)$.

Due to the labour involved in the computations, an extensive investigation of the spectrum for an arbitrary frequency modulation has not been undertaken. The general nature of the spectrum can be deduced from the radiation patterns of shaped-beam aerials.[7] The mean value of the spectrum will be given by equation (36), and ripples of the order of $\pm 20\%$ in amplitude and ± 15 deg in phase can be expected.

13. Appendix 4

The Received Spectrum and Pulse Shape

13.1. *Spectrum before Equalization*

The spectrum received from a target at angle θ_m is simply the transmitter spectrum (equation (12)) multiplied by the square of the aerial spectrum (equation (1)), specifically

$$G_1(f)=s(f).\{E(\sin\theta_m,f)\}^2 \qquad \ldots\ldots(44)$$

13.2. *Spectrum and Pulse Shape after Equalization*

The equalizer delay has the same slope as the transmitter delay $\tau(f)$, i.e. its characteristic is

$$\tau_e(f)=\tau_0-\tau(f) \qquad \ldots\ldots(45)$$

where τ_0 is any fixed delay. The variable delay term cancels the delay of the transmitted spectrum so that the output spectrum after equalization is

$$G(f)=A(f).\exp(-\mathrm{j}2\pi f\tau_0).\{E(\sin\theta_m,f)\}^2 \ldots(46)$$

The output to the 'all-beam' display is the Fourier transform of this spectrum, namely

$$H(t)=\int_{-\infty}^{\infty}G(f)\exp(\mathrm{j}2\pi ft)\,\mathrm{d}f \qquad \ldots\ldots(47)$$

Now the aerial spectrum $E(\sin\theta_m, f)$ is centred round a particular frequency f_m given by equation (6). If the amplitude of the transmitted spectrum $A(f)$ is reasonably constant over a band of frequencies (say about $4/D(f_m)$) centred on f_m, this term may be taken outside the integral, yielding

$$H(t) \simeq A(f_m)\int_{-\infty}^{\infty}\{E(\sin\theta_m,f)\}^2\exp\{\mathrm{j}2\pi f(t-\tau_0)\}\,\mathrm{d}f \qquad \ldots\ldots(48)$$

The output can now be determined from the convolution theorem of Fourier transforms. If we define

$$h(t)=\int_{-\infty}^{\infty}E(\sin\theta_m,f).\exp(\mathrm{j}2\pi ft)\,\mathrm{d}f \qquad \ldots\ldots(49)$$

so that $E(\sin\theta_m, f)$ is the inverse Fourier transform:

$$E(\sin\theta_m,f)=\int_{-\infty}^{\infty}h(t)\exp(-\mathrm{j}2\pi ft)\,\mathrm{d}t \qquad \ldots\ldots(50)$$

then the output becomes

$$H(t) \simeq A(f_m)\int_{-\infty}^{\infty}h(y).h(t-\tau_0-y)\,\mathrm{d}y \qquad \ldots\ldots(51)$$

The function $h(t)$ is related to the aperture illumination $g(x)$. Reverting to equations (1) and (2), the expression for the phase shift along the serpentine may be expanded in terms of the phase shift at the centre frequency f_m:

$$\psi(f) = \psi(f_m) - 2\pi(f-f_m)\times \times\left\{D(f_m) + \frac{(f-f_m)}{2!}\,.\,D'(f_m) + \ldots\right\} \quad (52)$$

If the aerial delay is reasonably constant over a band of frequencies centred around f_m, specifically if

$$D'(f_m) \ll \{D(f_m)\}^2 \qquad \ldots\ldots(53)$$

only the first two terms in equation (51) need be considered. The expression for the radiation pattern (equation 1) may then be re-written after some manipulation as

$$E(\sin\theta_m, f) \simeq \exp\left(-\mathrm{j}\frac{2\pi f_m R}{c}\right)\int_0^a g(x)\exp\left\{\mathrm{j}\frac{2\pi(f_m-f)R}{c}\right\}\times \times\exp\left\{\mathrm{j}2\pi x(f_m-f)\left[\frac{D(f_m)}{a} - \frac{\sin\theta}{c}\right]\right\}\mathrm{d}x \qquad \ldots\ldots(54)$$

Comparison with equation (50) yields

$$h(t) \simeq \frac{g(x)\exp\{\mathrm{j}2\pi f_m(t-R/c)\}}{\dfrac{D(f_m)}{a} - \dfrac{\sin\theta_m}{c}} \qquad \ldots\ldots(55)$$

where

$$x = \frac{(t-R/c)}{\dfrac{D(f_m)}{a} - \dfrac{\sin\theta_m}{c}} \qquad \ldots\ldots(56)$$

As a final approximation, we neglect the term $\sin\theta_m/c$ obtaining

$$h(t) \simeq \frac{a}{D(f_m)}\,.\,g\left\{\frac{a(t-R/c)}{D(f_m)}\right\}\,.\,\exp\{\mathrm{j}2\pi f_m(t-R/c)\} \quad (57)$$

The function $h(t)$ thus represents a pulse of carrier frequency f_m with an envelope shaped according to the aperture illumination of the aerial. Since $g(x)$ is zero outside the interval 0 to a, the duration of $h(t)$ is $D(f_m)$. It will be recognized that $h(t)$ defines the shape of the pulse which would have been received at the target if the transmitter spectrum had been equalized.

The shape of the output pulse may now be obtained from equation (51). For the particular case of a uniformly illuminated array:

$$g(x) = 1, \qquad 0 < x < a \qquad \ldots\ldots(58)$$

and

$$h(t) \simeq \frac{a}{D(f_m)}\,.\,\exp\{\mathrm{j}2\pi f_m(t-R/c)\}, \qquad R/c < t < \{D(f_m)+R/c\} \quad (59)$$

and the output pulse becomes:

$$H(t) = \frac{a^2A(f_m)}{D}\,.\,\exp(\mathrm{j}2\pi f_m t')\,.\begin{cases} 0, & t' < 0 \\ \dfrac{t'}{D}, & 0 < t' < D \\ \dfrac{2D-t'}{D}, & D < t' < 2D \\ 0, & 2D < t' \end{cases} \qquad \ldots\ldots(60)$$

where $$t' = t - \left(\tau_0 + \frac{2R}{c}\right) \qquad \ldots\ldots(61)$$

The output is thus a triangular-shaped pulse of carrier frequency f_m. The pulse duration is $2D$ and the peak occurs at $t = \left(\dfrac{2R}{c}+\tau_0+D\right)$, i.e. the delay is simply equal to the sum of the path delay, the fixed delay in the equalizer and the one-way aerial delay.

For tapered aperture illuminations, the output pulse shape is more heavily tapered. The duration is always $2D$ and the carrier frequency is always f_m.

13.3. *Effect of I.F. Filtering*

The pulse shape obtained above is valid for the output to the all-beam display For the individual beam filters shown in Fig. 1, the pulse shape will be modified further.

If the filter characteristic is $F(f)$, then the output may be obtained by further application of the convolution integral. The filtered output is given by:

$$H_1(t) = \int_{-\infty}^{\infty} H(y).V(t-y)\,\mathrm{d}y \qquad \ldots\ldots(62)$$

where $V(t)$ is the Fourier Transform of $F(f)$, namely

$$V(t) = \int_{-\infty}^{\infty} F(f)\exp(\mathrm{j}2\pi ft)\,\mathrm{d} \qquad \ldots\ldots(63)$$

$V(t)$ may be identified with the response of the filter to a Dirac function $\delta(t)$.

For example, a single tuned circuit of bandwidth B and centre frequency f_c has the characteristic:

$$F(f) = \frac{1}{1+\mathrm{j}\dfrac{2(f-f_c)}{B}} \qquad \ldots\ldots(64)$$

and the corresponding impulse response is

$$V(t) = \pi B.\exp(-\pi Bt).\exp(\mathrm{j}2\pi f_c t), \qquad t > 0 \ldots(65)$$

Using equation (60) to represent the filter input, the output may be evaluated from equation (62) yielding

$$H_1(t) = \frac{a^2 A(f_m)}{D} \cdot \left(\frac{\pi BD}{z_0^2}\right) \cdot \exp(\mathrm{j}2\pi f_m t') \times \begin{cases} 0, & t' < 0 \\ [(z-1)+\exp(-z)], & 0 < t' < D \\ [(1+2z_0-z)+\exp(-z)- \\ \quad -2\exp(z_0-z)], & D < t' < 2D \\ \exp(-z).[\exp z_0 - 1]^2, & 2D < t' \end{cases} \quad \ldots\ldots(66)$$

where

$$z = (\pi B t') \cdot \left\{1 + \mathrm{j}\frac{2(f_m - f_c)}{B}\right\} \quad \ldots\ldots(67)$$

$$z_0 = (\pi BD) \cdot \left\{1 + \mathrm{j}\frac{2(f_m - f_c)}{B}\right\} \quad \ldots\ldots(68)$$

and

$$t' = t - \left(\tau_0 + \frac{2R}{c}\right) \quad \ldots\ldots(69)$$

This is illustrated in Fig. 12 for the cases of a centrally tuned filter and one tuned to the -3 dB point of the target spectrum. In both cases, the filter bandwidth is assumed to be equal to the reciprocal of the aerial delay.

13.4. *Effects of Fluctuations in the Transmitter Spectrum*

One method of assessing the effects of fluctuations is to analyse them in the form of a complex Fourier series so that each term represents a spectrum of uniform amplitude and linear phase shift (i.e. constant delay). Extensive numerical work is often required.

In the particular case of a linear frequency modulation, Section 12.2 shows how the spectrum can be broken down into three main components, given by equations (29) (33) and (34). If the aerial spectrum is sufficiently removed from the edges of the transmitter spectrum, only the first term in the asymptotic expansions of equations (33) and (34) need be retained.

The delay in the main spectral component (equation (29)) is removed by the equalizer and the pulse is thus compressed to D. The subsidiary components (equations (33) and (34)) have substantially constant delays: hence the effect of the equalizer is to expand these 'pulses' to a length equal to the time taken for the transmitter to sweep through the band $1/D$. The duration of the subsidiary pulses is thus roughly $1/\alpha D$ and their peak amplitudes with respect to the main pulse are approximately

$$D(f).\sqrt{\alpha} \cdot \frac{|s_1(f)|}{|s_0(f)|} \simeq \frac{\alpha D(f)}{2\pi(f - f_1)} \quad \ldots\ldots(70)$$

and

$$D(f).\sqrt{\alpha} \cdot \frac{|s_2(f)|}{|s_0(f)|} \simeq \frac{\alpha D(f)}{2\pi(f_1 + \alpha T - f)} \quad \ldots\ldots(71)$$

The relative amplitude of each component is thus proportional to the rate of change of transmitter frequency and is inversely proportioned to the deviation of the target centre frequency from the edge frequency.

14. Appendix 5: Precipitation Clutter

14.1. *Echoing Area of Rain*

The equivalent echoing area of rain clutter which fills the beam at a particular range R is given by[14]

$$\sigma_R = \sigma_u \cdot \frac{c\zeta}{2} \cdot \frac{\pi}{4} \Theta \Phi R^2 \quad \ldots\ldots(72)$$

If the rain does not fill the whole of the elevation beamwidth, the value of Θ in equation (72) must be replaced by the angle subtended by the rain. For the particular case of a fairly broad beam with its lower edge at zero elevation, the subtended angle is

$$\frac{h_1}{R} - \frac{R}{2\rho}$$

where h_1 is the rain ceiling and ρ is the effective radius of curvature of the earth. The echoing area becomes:

$$\sigma_R = \sigma_u \cdot \frac{c\zeta}{2} \cdot \frac{\pi}{4} \Phi \left(h_1 R - \frac{R^3}{2\rho}\right), \quad \Theta > \left(\frac{h_1}{R} - \frac{R}{2\rho}\right) \quad (73)$$

Differentiation of equation (73) shows that the maximum echoing area occurs at a range

$$R_1 = \sqrt{\frac{2\rho h_1}{3}} \quad \ldots\ldots(74)$$

and the maximum echoing area is independent of the elevation beamwidth and is given by

$$\sigma_{R_{\max}} = \sigma_u \frac{c\zeta}{2} \frac{\pi}{4} \Phi \left\{\frac{8\rho h_1^3}{27}\right\}^{\frac{1}{2}}, \quad \Theta > \sqrt{\frac{2h_1}{3\rho}} \quad (75)$$

If the beamwidth is less than $\sqrt{2h_1/3\rho}$, the maximum echoing area occurs at a range where the beam is just filled, i.e. at the value of R given by

$$R_1 = \rho\left\{\sqrt{\Theta^2 + \frac{2h_1}{\rho}} - \Theta\right\} \quad \ldots\ldots(76)$$

The corresponding value of maximum echoing area is

$$\sigma_{R_{\max}} = \sigma_u \frac{c\zeta}{2} \frac{\pi}{4} \Phi . \Theta . \rho^2 \left\{\sqrt{\Theta^2 + \frac{2h_1}{\rho}} - \Theta\right\}^2, \quad \ldots \Theta < \sqrt{\frac{2h_1}{3\rho}} \quad (77)$$

As a particular example, assume:

$\Phi = \frac{1}{2}$ deg

$\sigma_u = 7 \times 10^{-9}$ (corresponding to 4 mm/hour at S-band[14])

$h_1 = 20\,000$ ft

$\rho = 8500$ km

then $R_1 = 185$ km

and $\sigma_{R_{max}} = 5{\cdot}4 \times 10^6 . \zeta$ $\Big\}\ \Theta > 1{\cdot}25$ deg .....(78)

whilst for $\Theta = 1$ deg,

$R_1 = 206$ km

and $\sigma_{R_{max}} = 5{\cdot}3 \times 10^6 . \zeta$ .....(79)

The maximum echoing area is thus about 5 m² for each microsecond of pulse length. A very narrow elevation beam, say less than 0·1 deg, is required if the maximum echoing area is to be reduced substantially.

14.2. *Calculation of Permissible Aerial Delay*

To simplify the argument, we neglect the effect of earth curvature and assume that the rain belt is of fixed altitude above the horizon, extending to the 'horizon' range $\sqrt{(2\rho h_1)}$. Specifically, we assume that the relationship between maximum range of rain and elevation angle is

$$R = \begin{cases} h_1 \operatorname{cosec} \theta, & \theta > \sqrt{\dfrac{h_1}{2\rho}} \\ \sqrt{2\rho h_1}, & 0 < \theta < \sqrt{\dfrac{h_1}{2\rho}} \end{cases} \quad \text{.....(80)}$$

We may then use equation (72) to compute the echoing area. This simplification gives some overestimate of the echoing area at low elevations. Using the radar parameters given in Section 4·1, and a value of 7×10^{-9} for σ_u, the echoing area becomes

$$\sigma_R = \begin{cases} 4{\cdot}7 \times 10^3 . \zeta \operatorname{cosec}^2 \theta, & \theta > 1{\cdot}1^\circ \\ 13 \times 10^6 . \zeta, & \theta < 1{\cdot}1^\circ \end{cases} \quad \text{...(81)}$$

If the receiver noise is not limiting, the target will be detectable when its echoing area exceeds the rain clutter by the minimum detectable signal, i.e. when

$$\sigma . L_p . L_s \geqslant S_{min} . \sigma_R . K \quad \text{.....(82)}$$

where L_p and L_s represent polarization and scanning loss factors applicable to the target and K represents the polarization loss or cancellation ratio applicable to the rain (L_p, L_s and K are all less than unity). Receiver noise will generally be negligible for the higher elevation conditions, but may begin to be troublesome in the long-range, low-elevation region. However, the overestimate of rain echo implicit in equation (81) will compensate for this. Using the radar parameters of Section 4.1, expressions for the compressed pulse length or aerial delay become:

$$\zeta \leqslant \begin{cases} 2{\cdot}5 \times 10^{-4} \sin^2 \theta, & \theta > 1{\cdot}1^\circ \\ 9 \times 10^{-8}, & \theta < 1{\cdot}1^\circ \end{cases} \quad \text{.....(83)}$$

The maximum permissible aerial delay is thus about 0·1 μs at low elevations, and increases to 30 μs at 20 deg elevation.

Manuscript received by the Institution on 2nd April 1964. (Paper No. 920/RNA29.)

DISCUSSION

Under the Chairmanship of Dr. M. I. Skolnik

Mr. M. A. V. Matthews: Figure 2 of the paper suggests that returns from different elevation angles θ_1, θ_2 are separated in time. How is this time separation distinguished from the time separation due to difference of range?

Dr. D. H. Davies: Would Dr. Milne comment upon the elevation finding accuracy of the new type radar relative to that of a conventional stacked beam radar?

The technique proposed by C. E. Cook and J. Paolillo† is a means of reducing the spectral ripples responsible for the side-lobes in the 'all-beam' output. Has Dr. Milne taken this technique into consideration?

† C. E. Cook and J. Paolillo, "A pulse compression predistortion function for efficient side-lobe reduction in a high-power radar", *Proc. Inst. Elect. Electronics Engrs*, **52**, No. 4, p. 377, April 1964.

Dr. Milne (*in reply*): The two targets shown in Fig. 2(g) are at different ranges. The pulses from targets at the same range but different elevations are coincident in time and the elevation information is contained in the carrier frequency of the pulse.

The elevation-finding accuracy of the radar depends upon the spectral response of the target. For the case of a point target where the echo amplitude is independent of frequency, the elevation accuracy will be the same as that of a conventional stacked-beam radar having the same beamwidth. In the case of aircraft targets, there is a need for considerable experimental work to establish the relationship between elevation accuracy, aircraft type, and aerial delay. The currently available experimental evidence is rather scarce, but shows that the echo amplitude has substantial correlation over a band of frequencies of about

10 Mc/s, corresponding to the reciprocal of the target length, but has little correlation over a band of 100 Mc/s. It thus appears that an aerial delay (and hence a compressed pulse length) of 0·1 μs will give little degradation in the elevation accuracy of an aircraft compared to a point target, but an aerial delay of 0·01 μs will show significant degradation.

I am grateful to Dr. Davies for pointing out the paper by Cook and Paolillo. This confirms that the fluctuations in the spectrum due to a linear frequency modulation and the resulting time side-lobes at $t = 0$ and T (Figs. 3 and 4 of the present paper) are due to switching the transmitter on and off. The technique of adding extra shaping to the frequency modulation at the beginning and end of the transmitter pulse to reduce these side-lobes requires an additional frequency change approximately equal to the original frequency change occurring during the pulse. For the present application, where the aerial selects only a small part of the transmitted spectrum, it is anticipated that the additional frequency change required would be comparatively small, and equal to the reciprocal of the aerial delay. Furthermore, it may only be necessary to introduce extra shaping at the start of the transmitter pulse, since the shaping already introduced to provide the cosecant-squared coverage diagram will tend to reduce the time side-lobe at $t = T$.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

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

July 1964	GBR 16kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	July 1964	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	− 150·7	− 150·6	0	17	− 150·1	− 151·2	+ 4
2	− 150·2	− 150·6	0	18	− 151·3	—	+ 1
3	− 150·2	− 150·6	0	19	− 151·3	− 151·3	+ 1
4	− 150·1	− 150·4	—	20	− 150·6	− 150·3	+ 2
5	− 149·7	− 149·6	—	21	− 150·2	− 150·0	+ 4
6	− 150·1	− 151·0	+ 1	22	− 149·9	− 150·1	+ 2
7	− 150·3	− 150·7	+ 3	23	− 150·3	− 150·6	+ 5
8	− 151·1	− 150·3	+ 1	24	− 150·1	− 150·7	+ 2
9	− 150·0	− 150·7	+ 3	25	− 150·5	− 150·6	+ 4
10	− 149·4	− 150·1	+ 3	26	− 151·1	− 152·1	+ 3
11	− 150·5	− 151·3	—	27	− 151·5	− 150·3	+ 2
12	− 150·9	− 152·1	—	28	− 149·9	− 150·5	+ 3
13	− 151·4	− 150·0	+ 4	29	− 149·4	− 150·9	+ 3
14	− 151·4	− 152·4	+ 4	30	− 150·7	− 152·2	+ 4
15	− 152·1	− 150·1	+ 5	31	− 151·1	− 150·9	+ 2
16	− 149·7	− 149·5	+ 3				

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

U.D.C. 621.317.72: 621.387.34

An Inexpensive Digital Voltmeter using Multiple-anode Dekatrons

By

P. GLEGHORN, B.Sc.(Eng.)†

Reprinted from the Proceedings of the Symposium on "Cold Cathode Tubes and their Applications", held in Cambridge from 16th to 19th March 1964.

Summary: A medium accuracy digital voltmeter which is robust, reliable and inexpensive is described; particular reference is made to the decade counters, which need only be capable of counting at moderate speeds.

Various types of counters are discussed, and a comparison made of their costs. The conclusion is reached that the use of a multiple-anode dekatron counting tube in conjunction with a cold cathode number tube provides the most economic method of counting.

A detailed description is given of the counters used in the voltmeter, including the drive circuits, the coupling of the dekatron to the number tube, and the reset and carry arrangements.

1. Introduction

In the early stages of their development, digital voltmeters were expensive and provided only moderate accuracies. Manufacturers therefore tended to aim for higher accuracy instruments where the cost could be better justified to the customer. This trend has been maintained with the introduction of more accurate and more expensive instruments until quite recently. However, the market is now more competitive, and customers require instruments with specifications only slightly inferior to those of the earlier models, but at much lower prices. In addition, the potential of an accurate instrument at a low price far exceeds that of previous instruments.

This paper describes some of the techniques employed in making digital voltmeters less expensive without seriously detracting from their accuracies. One of these techniques is the provision of a staircase waveform as the reference with which the input signal is compared; this in turn permits the use of a multiple-anode dekatron scaler circuit which can be directly coupled to a number-tube display.

A digital voltmeter can be broadly divided into two sections; the analogue, which includes the amplifier and the comparator, and the digital, which includes the control circuits and the method of forming the digital representation of the analogue voltage. At the accuracy which is under consideration, i.e. 0·1% of full-scale deflection, the problems of making the analogue circuits less expensive can be overcome without great difficulty; however, careful consideration must be given to the digital circuits.

† Gloster Equipment Ltd., Hucclecote, Gloucester.

2. General Digital Voltmeter Principles

Two well-known techniques have been used in the past to convert an analogue voltage into its digital representation. One method compares a current derived from the voltage to be measured with a succession of reference currents. Each reference current is produced by connecting a precision resistor in series with a voltage reference source; the smallest current corresponds to the resolution of the voltmeter. On completion of the process of comparison, the incoming signal is represented by a combination of these reference currents; their state, either on or off, is the required digital representation in binary form.

The second method of conversion transforms the analogue voltage into an equivalent time interval and measures this time by counting pulses generated by a crystal controlled oscillator. The analogue input signal is amplified to the required voltage level by an input amplifier, the output of which is compared with a voltage ramp of accurately defined slope. Two comparators are used to control a gate through which the pulses are passed to the counter. The first comparator opens the gate when the ramp passes through earth potential, and the second comparator closes it again when the ramp is equal in amplitude to the output of the amplifier. In this way, a gate is open for a time which represents the input voltage.

Because the accuracy of the voltmeter depends on the accuracy to which this time interval can be measured, the reference frequency is derived from a crystal. It must be small and robust, and its frequency must be sufficiently high to allow the maximum number of counts to be recorded within an acceptable

digitization time. These considerations lead to the adoption of an oscillator frequency of 1 Mc/s, and a first-decade counter capable of counting at this speed.

Both of these methods have disadvantages. The precision current comparison method presents considerable difficulties when attempting to calibrate the instrument, as only a given set of precision resistors and one setting of their associated switches can be checked with a single calibrating voltage. The ramp technique requires a crystal-controlled oscillator for the accurate time measurement, and a high-frequency decade counter capable of counting at the frequency of the crystal. There is, however, a lower frequency limit below which the use of a crystal is inconvenient due to its size; this frequency is above that at which cold cathode tubes operate. These disadvantages are overcome by a third method developed for the Digimeter, which whilst basically remaining a ramp technique embodies features of the other method.

3. The Digimeter

The method of conversion in the Digimeter obviates the need for an accurately defined frequency reference, thus allowing the crystal to be eliminated. Providing the upper limit of acceptable digitization time is not exceeded, the frequency can be reduced to a value compatible with the use of cold cathode tubes. During the digitization period, the displayed reading changes rapidly; an instrument is irritating to use if this period is too long, and 0·3 seconds is considered to be the desirable maximum period. The maximum count of the instrument is 999, and the clock frequency can thus be as low as 4 kc/s. This reduction in frequency also allows a simpler comparator to be used than those which have to operate in less than 1 microsecond.

In the Digimeter method of conversion, the ramp is replaced by a staircase waveform in which each step is an accurately defined 10 mV in amplitude and one period of the clock-frequency oscillator in width. The comparator compares the output of the input amplifier with the staircase waveform, and the number of steps in the staircase up to the time of operation of the comparator is counted. This represents the number of 10 mV steps in the voltage at the output of the input amplifier. The time taken by the staircase to reach the point of equality is immaterial, the accuracy being dependent only on the amplitude of the steps; thus the oscillator can be a simple astable circuit with its associated frequency drift.

A schematic diagram of the voltmeter is shown in Fig. 1. The pulses from the oscillator are limited by a Zener diode to ensure amplitude stability. An associated variable attenuator allows accurate setting of their amplitude, and so varies the slope of the staircase; this acts as a calibration control.

The pulses are applied to two capacitors in series; their ratio is chosen so that the pulse amplitude is divided to provide 10 mV steps at C2. Diode MR1 prevents the discharge of C2 at the end of each pulse, and each successive pulse therefore increases the potential across C2 by a further 10 mV. The potential across C2 opposes the charging effect of the pulse applied to C1 and C2. This would result in a non-linear staircase, so diode MR2 is therefore included to restore the incoming pulses to the level already attained by C2. The potential across C2 is applied to the emitter of the comparator transistor via an emitter follower which prevents leakage of the charge from C2. During the quiescent period between measurements, the clamp transistor is saturated, thus discharging C2.

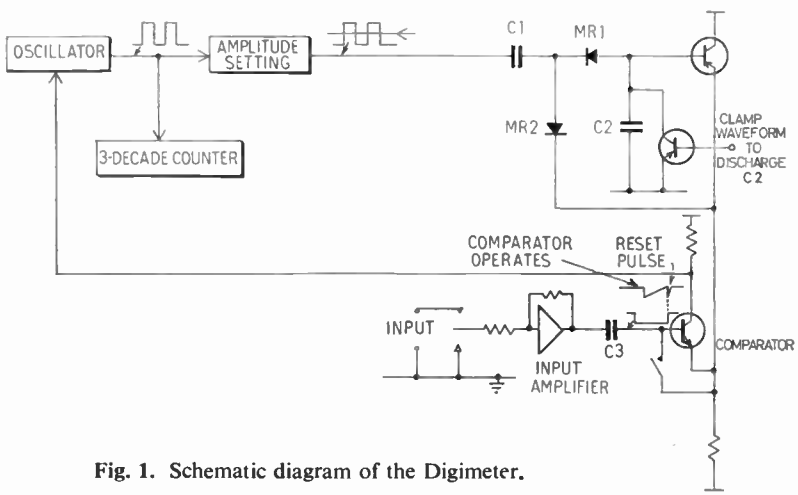


Fig. 1. Schematic diagram of the Digimeter.

The input amplifier is d.c.-coupled and has a feedback network around it providing either a gain of 1, which is used in conjunction with the input attenuator on 10, 100 and 1000 V ranges, or a gain of 10 to provide a 1 V range in which the least significant digit is 1 mV. The output of the amplifier must be referred to the starting voltage of the staircase, and must not include a component due to drift in the amplifier. Its input is switched between earth and the signal to be measured by a reed relay, and the output of the amplifier is coupled via capacitor C3 to the base of the comparator transistor. At this point the voltage waveform is a rectangular wave with amplitude proportional to the input signal, and is referred to the starting voltage of the staircase through another relay contact.

During the quiescent period, C2 is discharged by the clamp transistor VT1, and C3 charges to a potential determined by the drift in the amplifier and the voltage level at the start of the staircase waveform. At the end of the quiescent period, which occupies half the time between successive digitizations, the clamp is removed from C2, the counters are reset, and the relays de-energized. This applies a voltage step equal to the input signal amplitude to the input of the amplifier. This step is amplified by the appropriate factor and applied to the comparator.

After a short delay to allow for the rise-time of the amplifier and for the relays to settle, the oscillator applies pulses to the staircase waveform generator and to the counters simultaneously. When the two inputs to the comparator are equal in amplitude, the oscillator is clamped, and the displayed reading remains until the reset pulse at the commencement of the next digitization. At the beginning of the next quiescent period, VT1 again discharges C2, and the relays return to their original state, allowing C3 to charge to the new drift potential.

A specification of the counters on which the digital representation is held may now be derived. The period available for counting has to be limited to about 300 ms, a part of which is required for the amplifier output to rise to its final value and for the relays to settle. A count frequency of approximately 4 kc/s is therefore necessary to enable a count of 999 to be completed in the required time.

Other requirements the counter must meet are:

(1) to be capable of being reset with a single pulse;

(2) to provide a carry pulse to the succeeding decade;

(3) to provide a decimal read-out;

(4) to be capable of ready adaptation to provide print-out;

(5) to be inexpensive.

4. Types of Counter

Several different types of counter were considered for this application, and these are listed below. Some types were rejected in the early stages of the investigation, and in these cases the reason is indicated.

(1) Hard valve scalers; rejected on grounds of size and power requirements.

(2) Transistor scalers with projection display indicators.

(3) Transistor scalers with cold cathode number tube display.

(4) pnpn-diode ring counters; rejected as these diodes were expensive and at an early stage of their development when the decision on the type of scaler to be used was taken.

(5) Multi-aperture ferrite-core ring counters; rejected owing to difficulties of providing a read-out.

(6) Gas triode trigger-tube ring counters.

(7) Dekatron with inverting amplifiers to drive number-tube display.

(8) Multiple-anode dekatron directly coupled to number-tube display.

More detailed consideration is given below to these counters.

The method of providing a print-out facility is an adaptation of the counting operation, and is similar in principle in each case. This is considered in greater detail only in its application to the chosen method of counting.

The costs associated with each method are set out in Table 1.

Table 1

Comparison of costs of scalers

		Cost per Decade
Transistor Bistable Circuits	Binary-coded decimal with projection display	£31 1 0
	Binary-coded decimal with number tube	£28 16 0
	Ring counter with projection display	£42 15 0
	Ring counter with number tube	£39 17 0
Gas Triode	Ring Counter	£26 16 0
Dekatron	Conventional with inverting amplifiers	£28 5 0
	Multiple-anode with direct connection to number tube	£21 19 0

Allowance has been made in the prices to apportion the part of the power supply circuits and the resetting circuits which are required by each scaler.

4.1. *Transistor Scalers*

Conventional bistable circuits employing transistors can be connected in two ways to perform the operation of counting, either as a binary divider, with feedback to effect a division by 10 instead of by 16, or as a ring counter. The cheapest transistors function at considerably higher frequencies than the application requires, but the cost of the provision of the decimal read-out is high. In the first case, a decode matrix followed by buffer amplifiers must be provided; in the second case the connection to the display via buffer amplifiers is comparatively simple, but the cost of provision of the extra bistable circuits is high.

The scaler can be reset by a single pulse, but it has to be applied separately to each of the bistables. A carry pulse to the next decade can be obtained in certain configurations of the binary-coded decimal scaler from the pulse providing the feedback, but a separate gate to generate the carry is necessary in other arrangements. The ring counter provides a carry output direct.

The display used in conjunction with the transistor scaler can be either a projection display or a number tube. The former is capable of being viewed in strong light and is particularly attractive in being an in-plane display. It can be driven directly from the buffer amplifiers connected to the decoding matrix, but is expensive.

The number tube is a cheaper item, but requires transistors of higher working voltage as buffer amplifiers and the provision of a special power supply for its anode voltage.

4.2. *Gas Triode Scalers*

The gas triode trigger tube has a frequency range of operation compatible with the Digimeter requirement, and a low cost. However, the difficulties of a binary scaler with feedback are greater, as is the decoding of the output into decimal form, thus making the ring counter more attractive.

Certain complications arise as the circuits which circulate the 'on' state around the ring have to extinguish each triode as the 'on' state leaves it. Thus the simplest method of resetting the counter prior to digitization is to inject a negative-going pulse into the supply rail, and then to trigger the first tube in the ring of each decade. 'Carry' can be provided by using the transfer from the tenth tube to the first to advance the next decade; but similar difficulties are met in this instance to those encountered in circulating the 'on' state. The voltage levels in the gas triode scaler make direct connection to a number tube simple, and its power supply requirements are compatible with those of the gas triodes.

4.3. *Dekatron Scalers*

The use of a dekatron is particularly attractive in the Digimeter application as the space required is considerably less than the equivalent gas-triode circuit; the assembly work is also greatly reduced with the smaller number of components. All circuits associated with the transfer of the 'on' state from one cathode to the next are common to each decade counter further reducing the number of components required.

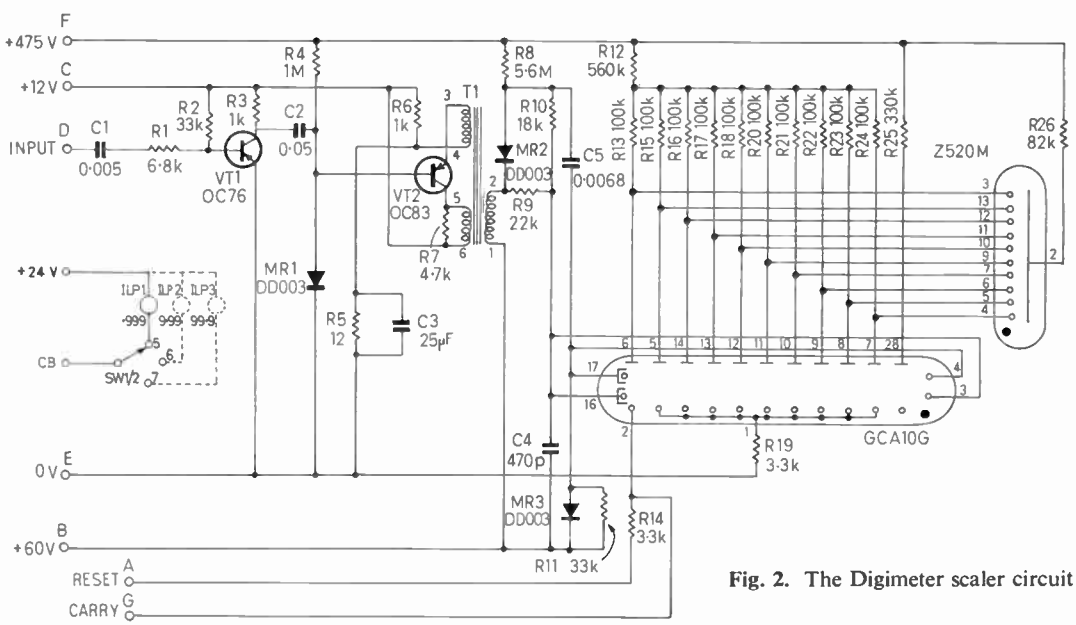


Fig. 2. The Digimeter scaler circuit diagram.

Propagation of the 'carry' pulse to the succeeding decade is particularly simple in the case of the dekatron circuit; a resistor can be inserted in the connection to the 'O' cathode and the voltage developed across this used as the input to the next decade. These resistors from the three decades are commoned together and a transformer inserted between this point and earth to inject the resetting pulse.

As the voltage waveforms at the individual cathodes of the dekatron are the inverse of those required by the number tube, direct connection between the two is not possible. An individual amplifier for each cathode would therefore be necessary to drive the number tube. These amplifiers would be gas triode trigger tubes or high voltage working transistors, but their inclusion would largely nullify the other advantages obtained by using dekatrons.

However, an alternative to the conventional dekatron has recently been introduced on the market, and overcomes the objection to the conventional dekatron scaler. The new dekatron is identical in its operation to the conventional type, but the anode is split in its construction into ten separate anodes, each of which has a separate connection. The cathodes, with the exception of the 'O' cathode, have a common connection. Using this multiple-anode dekatron, direct connection can be made from the anodes to the number tube cathodes; this provides a very simple and elegant method of obtaining a dekatron scaler with read-out.

5. The Digimeter Scaler Circuit

The circuit used is shown in Fig. 2. The anode load is partly common to all anodes, and partly exclusive to each individual anode. This prevents too great a potential occurring between the non-conducting anodes of the dekatron and the cathode, which might cause a discharge across the wrong path. Each anode is directly connected to a cathode of the number tube; the voltage waveforms applied to the associated cathodes ensure that the discharge in the number tube moves in sympathy with the circulation of the dekatron.

A small resistor is inserted in the common cathode lead and a similar one in the 'O' cathode lead. A potential of about 10 V is developed across this when the discharge is on the 'O' cathode; this provides the carry pulse to the following decade. The other end of this resistor is taken to the reset transformer, which is common to all three counters and is mounted externally to the counter boards. The reset pulse is injected at this point, causing the potential between the 'O' cathode and its associated anode to be much greater than that between the other cathodes, thus forcing the discharge to this path.

The driving circuits associated with the multiple-anode dekatron are similar to those used with conventional dekatrons. Since the drive to the first decade is at a different voltage level to the other two, being derived from an astable circuit providing a 0 to -12 V output, the input is a.c.-coupled to the first amplifier stage VT1. The output of the amplifier is a negative-going pulse of between 125 μs and 25 ms duration, according to the position of the decade in the counting chain. This is applied to the base of the blocking oscillator transistor VT2, and d.c. restored by the combination of MR1, R4 and C2. The leading edge of the pulse fires the blocking oscillator, which produces a negative pulse of 120 μs duration and 300 V amplitude across the tertiary winding.

This pulse is used to provide the drive waveform to the two guides of the dekatron; their waveforms are shown in Fig. 3. The drive to the first guide is applied via a differentiating circuit C5 and R11; diode MR3 removes any positive overshoot. The second guide

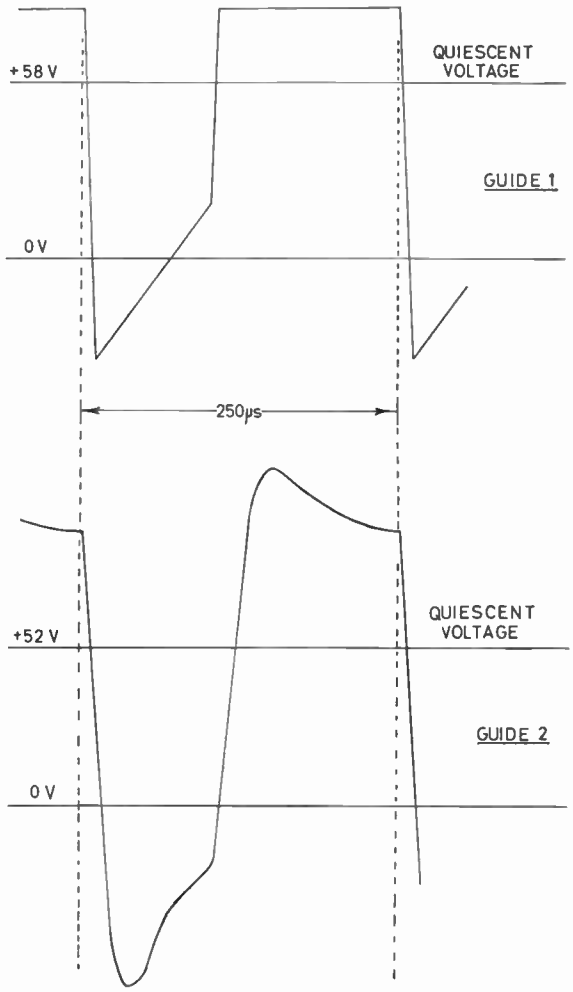


Fig. 3. Drive waveforms applied to the dekatron guides.

is supplied via an integrating circuit R10 and C4. The diode MR2 normally prevents any overshoot of voltage at the tertiary winding being transmitted to the guides, but the inclusion of R9 deliberately introduces a small amount of this overshoot to effect some improvement in the rise-time of the trailing edge of the waveforms applied to the second guide. This has the effect of producing a pulse whose amplitude is constant and independent of frequency.

The positive bias voltage must be carefully controlled for correct operation of the counters. If the bias is too low, the operation will remain unaffected but the life of the dekatron will be reduced; too high a bias reduces the effect of the negative-going portion of the drive waveform, and causes the intermittent loss of counts. At one stage of the development, the windings on the transformer had been revised from the original version in order to use a different bobbin, and circuit tolerances could permit the bias rail to be at a voltage which would cause intermittent operation. This was corrected before the production stage by increasing the number of turns on the tertiary to provide a greater output amplitude. A check of the available safety margin was made by plotting a graph of bias voltage against the number of lost counts; this enabled the point at which intermittent operation commenced to be accurately determined, and the required output of the tertiary winding to be set.

6. Print-out

A method of print-out was required which would be compatible in price with the Digimeter. However, printers in the lower price bracket require a serial pulse input at about 10 impulses per second; some form of buffer circuit must therefore be used to connect the Digimeter to the printer.

The print-out circuit is not shown on the scaler circuit diagram, but its principles are illustrated in Fig. 4. The printer has a printing drum which is rotated by a solenoid-operated ratchet mechanism; each operation advances the drum by one digit. After the required number of solenoid operations, a print command is given which impresses the drums on the paper. To set these drums to the required positions, the normal interconnections of the three decades are removed; a train of ten pulses is then injected simultaneously into each decade, circulating the dekatrons once and returning them to their original positions. The train of ten pulses is also applied to the drum solenoids of the printer via a gate, but the gate is only opened after the dekatron has passed the 'O' cathode; thus the number of digits accumulated by the drum after the circulation of the dekatrons is equal to the number displayed by the number tube. A print command pulse is then given to the printer.

The advantages of the dekatron scaler when used in this printing application are that a ready indication of the 'O' cathode may be obtained from the same point as the carry pulse is generated; also very little modification of the basic counter circuit is required to incorporate the additional 'print' components.

7. Conclusions

This paper describes how the adoption of certain techniques in digital voltmeter design enables a pulse source to be used which does not require the accuracy of a crystal. The frequency of the pulse source can therefore be decided by other factors, and does not have to be in the band in which small robust crystals are available. In this particular application where digitization time and the total number of

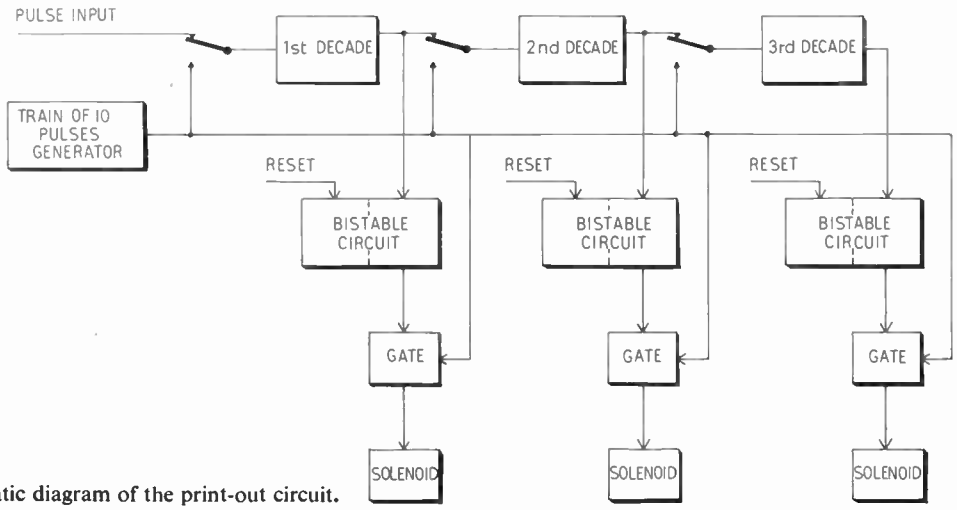


Fig. 4. Schematic diagram of the print-out circuit.

counts permit, the oscillator frequency can be within the frequency spectrum of cold cathode devices.

The survey of all suitable counting methods available at the time of the Digimeter development shows that the most economic approach to the requirement is provided by a scaler using a multiple-anode dekatron with direct connection to a number tube. Facilities for reset are readily available in this scaler, and it can be simply adapted for connection to print-out circuits.

8. Acknowledgments

The author wishes to thank the directors of Gloster Equipment Ltd. for permission to publish this paper, the Applications Department of Ericsson Telephones Ltd. for the original counter circuit, and Mr. D. May of Gloster Equipment Ltd. for his early work on the instrument.

Manuscript received by the Institution on 10th January 1964, (Paper No. 921.)

DISCUSSION

Under the Chairmanship of Mr. D. Reaney

Mr. A. G. Wray: I note that the generation of the 10 mV step involves a 1000 : 1 capacitance divider feeding into the base of transistor. It is essential for accurate functioning of the circuit that the generated waveform shall have a perfectly flat top, i.e. that no differentiation should take place.

Would the author please indicate the size of capacitances used in the divider network and comment upon the influence of stray resistance (as for example the input resistance of the transistor following the divider network) on the performance of the equipment.

The Author (*in reply*): The instrument is calibrated against an internal Zener reference at both ends of the scale, and hence the slight discharge of the capacitor across which the staircase waveform is generated can be taken into account. This assumes that the clock frequency remains substantially constant, but errors due to these variations can be regarded as a second order effect. The staircase is developed across a 4 μF capacitor, and the following transistor is a compound emitter follower; this ensures a time constant in excess of ten times the digitization time of the instrument.

Dr. C. G. Morgan: Is there a minimum duration of input voltage impulse below which the digital voltmeter you described will not operate? For example, if the voltmeter were fed with a single rectangular impulse of 0·1 ms duration, would the voltmeter accurately display this voltage?

The Author (*in reply*): A small capacitor is connected across the input of the amplifier to reduce the effects of any contact bounce on the input relay; in order to measure a pulse accurately its duration must be sufficient to charge this capacitor through the input circuit. This means that the pulse must be applied for a minimum of 10 ms before the relay changes.

INSTITUTION NOTICES

Standards and Specifications Sub-committee

The Institution has supported the British Standards Institution for many years and is represented on the Telecommunication Industry Standards Committee and on the appropriate Technical Committees of B.S.I. Members serving on these Committees have from time to time been called upon to represent the British National Standards Committee of the International Electrotechnical Commission at I.E.C. meetings.

The Institution's Technical Committee has been charged by the Council with co-ordinating this activity and with making recommendations for representatives to serve on B.S.I. Technical Committees. However, in recent years standards matters have become much more prominent and it has become necessary to form a sub-committee to deal with this work on behalf of the Technical Committee. A Standards and Specifications Sub-committee has been set up with the following terms of reference:

1. To make proposals and recommendations for the formulation of standards, specifications and codes of practice on behalf of the Technical Committee.
2. To co-ordinate the activities of representatives serving on technical committees of the British Standards Institution and to suggest new representatives as required.
3. To suggest committees on which representation should be sought.
4. To deal with draft national and international standards documents submitted for comment.
5. To deal with other matters referred to the Sub-committee by the Technical Committee and to propose to that committee any other matters worthy of consideration.

The Sub-committee is under the chairmanship of Mr. G. R. Jessop (Associate Member) and includes the following members: Mr. F. G. Diver, Wing Commander F. C. Lowe, Mr. G. Hersee and Mr. J. M. Peters (Members) and Mr. C. J. Hansford (Associate Member).

Current Papers

Readers of the June issue of *The Radio and Electronic Engineer* and of the July issue of *IEE News*, which were published within a day or two of each other, will have seen that, coincidentally, both the I.E.E. and I.E.R.E. pointed to the need for a new publication, produced quickly and cheaply, that would give the titles and references of current literature in the field of electrical and electronic engineering. Both Institutions stated that they were considering providing such a service. It would clearly have been wasteful duplication for both to proceed. The I.E.E. already published *Science Abstracts*, which entails the acquisition and scrutiny of nearly a thousand periodicals. They were thus more favourably placed to meet the new requirement and have accordingly undertaken the work.

The first issues of the new publication, entitled *Current Papers*, are being distributed to all members of both Institutions with an invitation to subscribe, and it is hoped that many members will do so. All subscriptions should be sent direct to the I.E.E.

This announcement is being published in the *Journals* of both Institutions.

Conference on "Lasers and their Applications"

The I.E.E. Electronics and Science Divisions in collaboration with the I.E.R.E. and the United Kingdom and Eire Section of the I.E.E.E. are organizing a Conference on "Lasers and their Applications" to be held at the Headquarters of the I.E.E. from 29th September to 1st October 1964. The conference will be divided into seven sessions covering the following broad subjects:

Gas lasers;

Solid state and injection lasers;

Modulation, demodulation and detection;

Propagation and communications;

Measurement techniques and radiation hazards;

Ranging, navigation and meteorology;

Machining and welding—uses in physics and medicine.

Further details and application forms may be obtained from the Institution at 8–9 Bedford Square, London, W.C.1.

U.K.A.C. Annual Lecture

The Fourth Annual Lecture of the United Kingdom Automation Council will be given by Mr. W. F. Cartwright, D.L., J.P., M.I.Mech.E., on Tuesday, 6th October 1964, at the Institution of Electrical Engineers, the subject of this lecture, which begins at 5.30 p.m., will be "The Future of Automation in the Iron and Steel Industry". Admission to the lecture is free but tickets should be obtained in advance from the Honorary Secretary, U.K.A.C., c/o The I.E.E., Savoy Place, London, W.C.2.

Index to Volume 27

The June 1964 issue completed Volume 27 of *The Radio and Electronic Engineer* which covers the period January–June 1964. An Index to the volume is enclosed with this issue.

U.D.C. 621.385.832.2

Picture Tubes for Television Displays using Quadrupole Scan Magnification

By

K. E. JOHNSON, B.A.
(Graduate)†

Summary: The design and performance of picture tubes having 8½-in screens, 90 deg deflection angle, and using permanent-magnet quadrupole scan magnifying lenses is described. These tubes require only a thirtieth or less of the line scanning power of a conventional tube having the same final anode voltage and deflection angle. Field scanning power is also slightly reduced.

Descriptions of both magnetic and electrostatic focusing systems are given. The magnetically focused tube suffers from the disadvantages that it requires an extremely stable e.h.t. supply and a large number of focusing adjustments. The electrostatically focused tube requires no more stabilization of e.h.t. than is usually provided for conventional tubes, and reduces the number of focusing adjustments considerably.

However, the increased length of the tube (40% greater than conventional tubes) and increased complexity of manufacture and adjustment compared with conventional tubes, coupled with the need for a separate e.h.t. and video h.t. supply generator, has led the author to the conclusion that this technique has little to offer at the present stage of picture tube development.

1. Introduction

In the past few years the problems associated with using transistors in television sets have received a great deal of attention. Not the least of these problems are the time-base circuits, in particular the line time-base. This circuit absorbs greater power (up to two-thirds of the total) than any other circuit in the set. The specification for the output transistor to drive the coils is also an important consideration. This transistor must be capable of handling high peak collector currents while also being able to withstand the peak voltage occurring between collector and emitter during the flyback period.

Any method of reducing the required energy to scan a typical 90 deg or 110 deg picture tube is therefore of interest, since the total power requirement of a receiver may be considerably reduced, while enabling cheaper transistors to be used in the output stage of the line time-base circuit. The battery-driven receiver is an example where such a method could be of great value in contributing to maximum battery life. The tubes described in this paper were designed with the aim of reducing the required scanning power by the use of scan magnification techniques.

Scan magnification is the name given to a general technique for increasing the deflection sensitivity of cathode-ray tubes. Basically, in the case of a television picture tube a static field placed between the coils and the screen is arranged to magnify an initially small deflection produced by the coils so that the whole screen is scanned. The strength and axial length of the field determine the magnification factor. In terms of an optical analogy, this field is equivalent to a diverging lens. Unfortunately, few types of diverging electron lens are suitable for this particular application. Much research has been carried out in the United States into the application of rotationally symmetrical electrostatic mesh lenses as scan magnifiers.[1, 2] In the Mullard Research Laboratories, however, work has been concentrated on the use of non-rotationally symmetrical quadrupole lenses. These offer the advantages for television purposes that beam current is not intercepted by mesh wires, so that in theory at least better picture contrast and brightness should be obtained with quadrupole scan magnifiers; and that larger magnifications of deflection are possible (in one plane) than with the mesh lens.

The principle of quadrupole scan magnification was briefly described in 1958[3] in connection with an experimental battery-driven transistorized television receiver which used a 17-in tube and magnetic quadrupole focusing and scan magnifying lenses. These were constructed from 'Ticonal' bar magnets mounted in machined light alloy holders. This first exercise demonstrated the defects as well as the advantages of scan magnification, and showed the necessity for further development.

† Mullard Research Laboratories, Redhill, Surrey.

Later work led to a completely portable transistor television receiver using an $8\frac{1}{2}$-in tube and permanent magnet 'Magnadur' quadrupoles (magnetized on special jigs), to reduce the cost of the lenses. Although this second exercise showed that cheaper magnetically-focused displays could be built successfully, it was decided to develop tubes incorporating electrostatic focusing to overcome the necessity for complicated mechanical adjustments on the magnetic focus unit, and complex circuitry to stabilize the e.h.t. Tubes using electrostatic focusing lenses and a magnetic scan magnifier were demonstrated in July 1962. A description of these tubes forms the main object of this paper.

For completeness a restatement of the method of scan magnification using quadrupole lenses and some general remarks on the choice of tube dimensions are given in the following sections. Next, details of practical scan magnifiers are given. A section on magnetic quadrupole focusing units follows as details of these have not previously been published. The development of electrostatic quadrupole focusing is described, and finally, a short assessment of the advantages and disadvantages of these tubes is given.

2. The Principles of Scan Magnification using Quadrupole Lenses

2.1. *The Scan Magnifier*

Figure 1 shows the lines of force in a cross-section of a typical magnetic scan magnifying quadrupole lens described in a later section. Consider the effect of this field on an electron beam travelling perpendicularly to the cross-section towards the reader. The beam will be diverged in the horizontal plane of the diagram, and converged in the vertical plane. The optical analogue of this is a 'saddleback' lens (Fig. 2). If a beam of an initially round cross-section enters the lens, the cross-section of the beam leaving the lens will be an ellipse with its major axis in the horizontal plane.

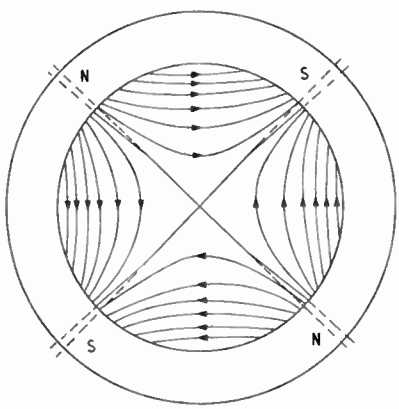


Fig. 1. Cross-section through field of scan magnifier lens.

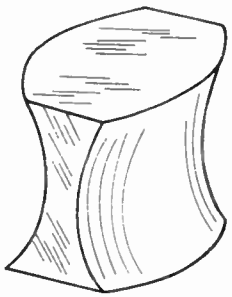
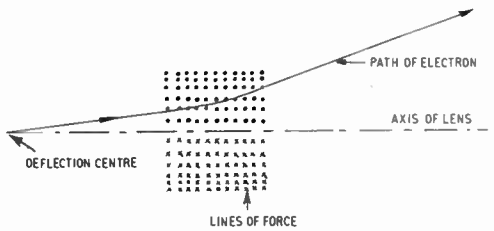
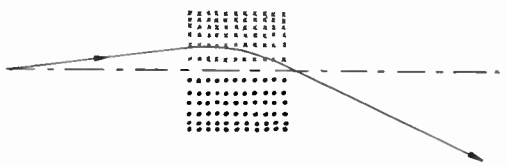


Fig. 2. Optical analogue of quadrupole lens.

The diverging action in one plane of this lens is just what is required of a scan magnifier. However, for small magnifications the converging action in the perpendicular plane will result in a decreased scan in this plane at the screen. This can be avoided by increasing the power of the lens to such a value that the incoming deflected beam in the vertical plane crosses the axis of the tube just after the lens. The scan is thus inverted. Figures 3(a) and 3(b) show cross-sections through the axial planes of 'mirror' symmetry of the field, and the paths of electrons travelling in these planes. The equations of these paths are derived in the Appendix, where it is shown that due to the inversion, a small increase of scan is possible in the vertical plane, together with a much enlarged scan in the horizontal plane.



(a) In horizontal plane of scan magnifier.



(b) In vertical plane of scan magnifier.

Fig. 3. Ray path.

2.2. *The Focusing System*

Conventional focusing methods cannot be used with this type of scan magnifier. In a typical cathode ray tube the spot on the screen is an image of the gun

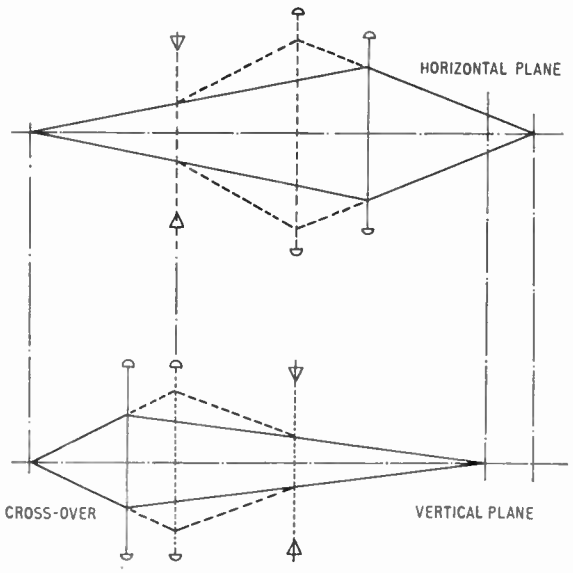


Fig. 4. Equivalence of crossed cylindrical lenses and crossed quadrupole lenses as prefocusing systems.

cross-over formed by the focus lens. It can be shown that to adopt the same method in a scan magnifying c.r.t. requires the replacement of the conventional focus lens by an unconventional non-rotationally symmetrical lens system.

Consider the beam shape entering the scan magnifier for the required converging beam of circular cross-section leaving it. The in-going beam must be converging, but not equally so in the vertical and horizontal planes. The effect must be to produce, if the scan magnifier field is zero, two mutually-perpendicular line images separated from one another along the axis by a small distance and both situated within the region occupied by the scan magnifier. The vertical line image in the case of the tubes to be described must be the nearer to the screen and approximately a sixth of the width of the horizontal line image.

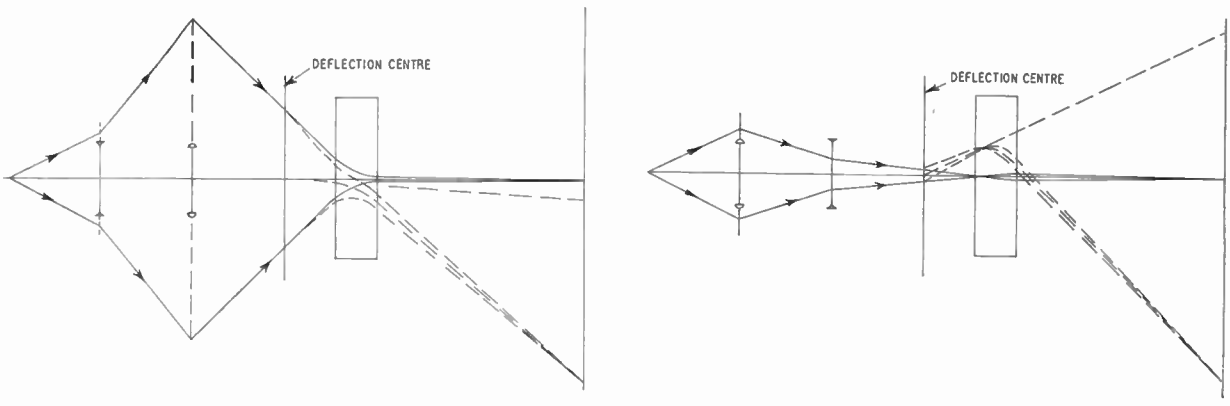
These line images of the round gun cross-over can be produced by two crossed cylindrical focusing fields as shown by the full lines in Fig. 4. These fields are, however, equivalent to the pair of crossed quadrupole fields shown by the dotted lines. The latter is the preferable practical arrangement since the position of the cylindrical focusing field producing the vertical line image is close to and may even coincide with the position of the deflector coils. The crossed quadrupole arrangement has therefore been adopted in the tubes to be described. Figures 5(a) and 5(b) show the two planes of the complete system in schematic form.

An interesting point in this context is the use of unconventional electron guns with elliptical grid holes in quadrupole focusing systems. These appear to offer some possibility of simplifying the above system but will not be discussed here.

3. Some General Remarks on the Design of Scan Magnifying Displays for Portable Receivers

Whereas the initial experiments on scan magnification were based on a 17-in tube,[3] later experiments were conducted using smaller tubes, with the possible application of the tubes in portable receivers in mind. In this section the general considerations which determined the choice of bulb size, etc., are given.

A portable television receiver must be as light and compact as possible for ease of carrying. The cathode-ray tube must be capable of adequate resolution and good contrast at high peak brightness to cope with high ambient lighting conditions. When the receiver is operated from its own or an external (e.g. car) battery supply, the current drain from the battery should be as low as possible for long battery life, and to allow the use of small light-weight batteries.



(a) In horizontal plane of scan magnifying system.

(b) In vertical plane of scan magnifying system.

Fig. 5. Schematic form of typical ray paths.

The first of these considerations limits the maximum permissible screen size, but this is an advantage when considering the high picture brightness required. A compromise has to be made between the conflicting requirements of size, weight and performance. The tubes to be discussed had a screen diagonal of 8·5 in and a deflection angle of 90 deg. (The bulbs used were identical with those of the AW21–11 television tube.) This deflection angle provided a suitable compromise between the requirements of minimum deflection power and a short bulb length.

The neck length and diameter were determined by several factors, and the best compromise has not necessarily been chosen. For a small cabinet depth, the overall length of the tube, and therefore the neck length should be as small as possible. In the interest of picture brightness a high beam current, wide beam angle, triode-type electron gun was used. The construction of this gun required a standard 36 mm diameter neck. Although a smaller neck diameter would have meant a decrease in the deflection power required, it was felt that this diameter would enable large aperture quadrupole lenses to be used, thus assisting in reducing lens aberrations to a minimum. Having adopted this neck diameter and with a nominal final anode potential of the gun of 18 kV the necessary scan magnifier power, and therefore length, was fixed for optimum magnification of scan. The remaining neck length was therefore determined by the deflector coil length, and the design of the crossed quadrupole focusing fields.

A further point which should be mentioned here concerns the resolution of scan magnifying tubes. Theory shows that when any scan magnifying tube is compared with a similar tube without scan magnification, the spot size is always greater in the scan magnifying tube by a factor larger than the magnification of scan. This effect may limit the maximum scan magnification factor which can be used, and in tubes using quadrupoles where the factor is fixed by the design of the scan magnifier, other steps may be necessary to limit the spot size.

4. Magnetic Scan Magnifiers

The design and method of magnetization of scan magnifying quadrupole lenses, and the method used to correct for the raster distortion associated with them is given in the following section. Early experiments with both magnetic and electrostatic scan magnifiers indicated that the development of the latter would be considerably more difficult than that of the former from a technological point of view, and this has not been attempted.

4.1. *Construction and Method of Magnetization*

The permanent magnet quadrupole lenses developed for these tubes were made from rings of magnetically-isotropic, sintered 'Magnadur 1'. Those for the magnetically-focused tube had nominal dimensions of 78 mm outer diameter, and 44 mm inner diameter, though these varied by 1 or 2% between different batches of rings. They were ground to the required thickness before being magnetized to saturation on a special jig. Their focal lengths were therefore determined for a given beam potential.

Small adjustments to their focusing properties could be made when necessary by using suitably shaped soft iron shunting pieces placed at appropriate positions on or near their end faces.

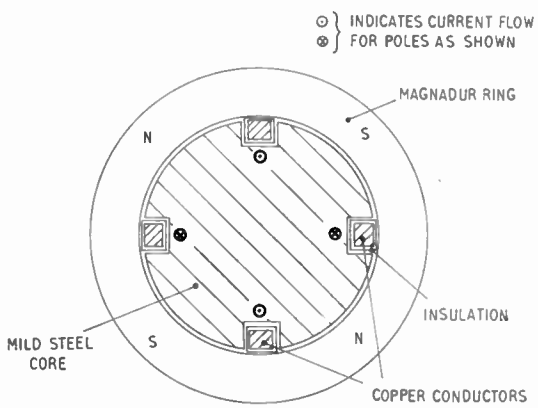


Fig. 6. Cross-section of typical magnetizing jig.

The cross-section of a typical magnetizing jig is shown in Fig. 6, with a ring in position on it. The diameter of the jig was such that a ring would just slip over it. Opposite copper conductors were connected in parallel, and the pairs were connected in series. For the directions of the current flow shown the effective poles appearing at the inner surface of the ring are marked. The jig was several inches long so that more than one ring could be magnetized at a time. The mild steel core provided a low reluctance path for the lines of force where they would otherwise have passed through air. The insulation pieces around the conductors were necessary to prevent shorting through the uninsulated steel core.

To ensure the symmetry of the field and obtain complete saturation of the material the method used was to pass a high current pulse through the jig; next, to turn the ring through 180 deg and repeat, and then to repeat this cycle several times.

Measurement of the field within the central region of the lenses was not easy because it was difficult to make sufficiently small accurate probes whose position in the field could be known with adequate precision. In spite of these problems, measurements using a saturation magnetometer were made. These showed

that the peak values of the field gradients $\partial H_x/\partial y$ and $\partial H_y/\partial x$ in the two symmetry planes of the lenses were in the region 450–550 gauss/cm near the axis.

From these results, the quadrupole constant λ derived in the Appendix and given by:

$$\lambda = \left(\frac{e}{2m\phi}\right)^{\frac{1}{2}} \left(\frac{\partial H_x}{\partial y}\right)^{\frac{1}{2}} = \left(\frac{e}{2m\phi}\right)^{\frac{1}{2}} \left(\frac{\partial H_y}{\partial x}\right)^{\frac{1}{2}}$$

where e/m is the ratio of charge to mass for an electron and ϕ is the beam potential, was calculated to be of the order of $0{\cdot}1\ \text{mm}^{-1}$ for an 18 kV beam.

With this value of λ, the effective length, L, of the scan magnifier for maximum vertical magnification must be about 23 mm. The actual thickness for this condition was determined empirically to be 18 mm. The fact that this is less than the calculated effective length of the lens is a consequence of the end effects of the field which are not taken into account in the calculation, and which are highly significant in a lens whose length is not large compared with its aperture. Measured scan magnification factors for the horizontal and vertical planes agreed fairly well with the values calculated in the Appendix, being about ten times in the line plane and just less than two in the field plane.

In an attempt to minimize the size and weight of the scan magnifier, lenses of the type just described but made from smaller rings have been developed. These were constructed from 'Magnadur' rings of outer diameter 55 mm, inner diameter 39 mm, and various thicknesses from 3 mm to 10 mm, the rings being stuck together to form lenses of the required strength. A lens of the same length as the previously described scan magnifier (18 mm) proved to be rather weaker, giving magnifications of approximately eight times in the line plane and unity in the field plane. However, a useful saving of 2·8 times in volume, and thus in weight, was obtained.

4.2. *The Correction of Raster Distortion*

In this system the deflection produced by the line coils is so small that any contribution to raster distortion can be ignored. The same is not true of the field coils; the deflection produced by them has to be practically the same as that produced by a conventional coil. Hence they will provide a contribution to the raster distortion of the same order as in a conventional tube. In the main, however, raster distortion is caused by the increased time spent in the scan magnifier field for an electron having both horizontal and vertical deflection.

Burfoot[4] has shown what types of distortion can be produced in systems having symmetry about two planes at right angles, of which the quadrupole scan magnifying system is an example. The usual pin-cushion and barrel distortions occur, together with two other distortions which cannot be distinguished practically from these two. However, two further possibilities arise in which, for example, the top and bottom edges of a picture may suffer from barrel distortion, while the two vertical edges are pin-cushion in shape. These Burfoot calls 'hammock' distortions.

With a lens magnetized as described above, about 13% pin-cushion distortion was observed in both vertical

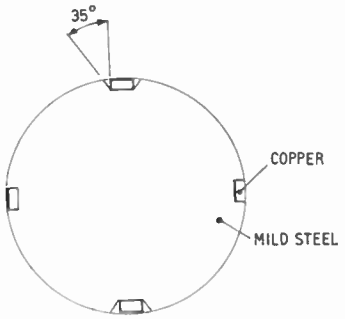
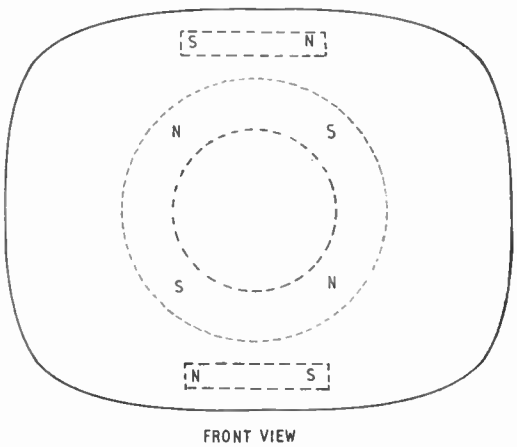


Fig. 7. Required modification of magnetizing jig to correct vertical edges of raster.



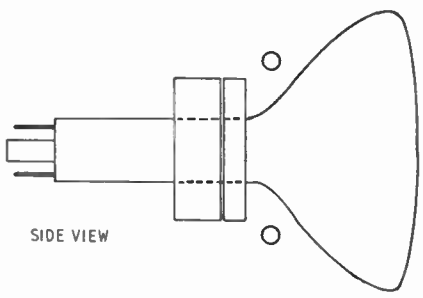
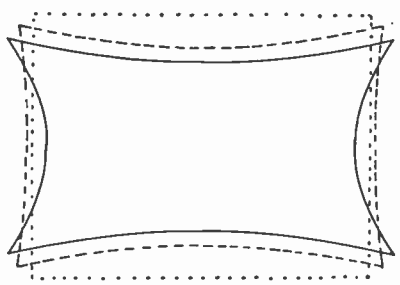


Fig. 8. Position of small bar magnets to correct horizontal edges of raster.

and horizontal raster edges. The method of correcting the vertical edges in this tube was to increase the field near the edge of the scan magnifier aperture in the horizontal plane, so that the middles of the edges were pulled out. This was accomplished by moving the effective pole centres of the lens towards the horizontal plane as shown in Fig. 1. Figure 7 shows the required modification to the previous magnetizing jig to do this, the amount of the mild steel core to be cut away having been found empirically. Small amounts were cut away until the raster produced by a lens magnetized on the jig was of the desired shape.

The 35-deg cut-away shown in Fig. 7 reduced the vertical edge distortion almost to zero, while leaving the horizontal edges practically unaffected. These edges were corrected by small bar magnets placed just after the scan magnifier close to the cone of the tube (Fig. 8). These also reduced the remaining small amount of vertical edge distortion to zero. Figure 9 shows schematically the stages in correcting the raster.



——— Uncorrected raster.

– – – – Effect of modifying the magnetizing jig.

. . . . Straight-edged raster produced by the combined effect of the modified jig and correcting magnets.

Fig. 9. Stages in correcting raster distortion.

In practice, although it was possible to correct the edges of the raster completely, in regions intermediate between the centre of the picture and the edges, combinations of distortions were produced. Other compromises were obtained which corrected these regions but left a small amount of distortion in the edges of the picture. Completely corrected pictures were not obtained, although satisfactory pictures were achieved. The positions of the two separate raster correcting magnets were critical and required careful adjustment to give rasters with a remaining distortion of less than 1 or 2%.

4.3. *Another Practical Approach to Raster Correction*

Mention should be made of another approach to the problem of raster correction in scan magnifying tubes of this type. The scan magnifier, instead of being a solid sintered 'Magnadur' ring, was made up from thin (of the order of 1 mm) sintered or bonded 'Magnadur' rings, magnetized on special jigs and arranged in such a way that separate raster correcting magnets on the cone of the tube were unnecessary. This was an extension of the idea described above in which the external regions of the quadrupole field were influenced to produce the desired stretching of the raster, so compensating for the pin-cushion distortion. Lenses were built up consisting partly of rings designed to straighten the vertical edges, and partly of rings to straighten the horizontal edges of the raster.

A lens was produced which gave a straight-edged raster, but it proved difficult to make copies of it due to the large dimensional and magnetic tolerances associated with this type of bonded ring. No further development has been carried out to overcome this difficulty.

5. A Magnetic Focus Unit

As discussed in Section 2, non-rotationally symmetrical focusing is required when using a quadrupole scan magnifier to compensate for the different focusing properties of the latter in the horizontal and vertical planes. This prefocus system can take the form of two crossed magnetic quadrupole rings (magnetized on the jig of Fig. 6) and separated by a small distance.

5.1. *Construction of the Unit*

The focal lengths and positions relative to screen and gun of the two quadrupoles forming the focus unit can be calculated, making the assumption that they are 'thin' lenses and providing that the deflector coil and scan magnifier parameters are known. For a separation from the end of the focus unit to the beginning of the scan magnifier of 80 mm, the calculated separation of lens centres is 25 mm, and the centre of the first lens to grid plane distance is 45 mm. The focal lengths are 23 mm and 29 mm for first and second lenses respectively. Unfortunately, the quadrupole constant determined previously for a single ring with 78 mm outer diameter and 44 mm inner diameter cannot now be used to calculate the length of each ring because of the shunting action one ring has on the other, thus reducing the field gradient near the axis. It was found by experiment that for two such rings whose centres were separated by about 25 mm, a length of 12 mm was necessary for the first ring to obtain the required focusing.

In an actual unit, shown mounted on a tube in Fig. 10, each ring was made 12 mm long, the slightly lower power of the second lens being obtained by shunting it with a thin mild steel ring whose position relative to the lens could be adjusted. (Note that this length indicates a reduction of field gradient near the axis, due to the shunting effect between the lenses, to about one-third of its unshunted value.)

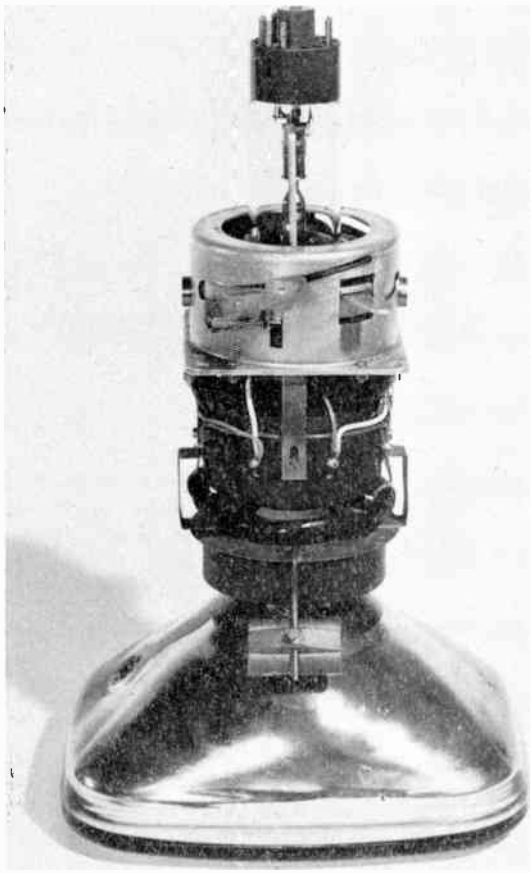


Fig. 10. Magnetically focused tube.

5.2. *Adjustments*

Because of the wide dimensional and magnetic tolerances of 'Magnadur' rings, considerable freedom of adjustment of the rings was necessary to obtain a focused spot. The necessary adjustments were:

(1) rotation of one ring relative to the other;

(2) vertical and horizontal movement of each ring;

(3) relative distance between rings;

(4) axial position.

These adjustments involved seven separate movements. The rings were mounted in an aluminium housing which was positioned by slides on the chassis of the receiver. The housing was arranged to give adjustments (1) and (3), while the slide gave (4); (2) was obtained by arranging a flexible mounting for the tube face, and having the base mounted in an adjustable plate.

Observations showed that in general the ellipticity of the spot could be controlled by the separation of the rings, while the overall spot size depended mainly on the mean position of the rings on the tube neck. Usually adjustment of these two quantities, together with the position of a shunting ring which varied the second lens power, was sufficient to allow for spreads in different pairs of rings.

5.3. *Correction of Aberrations due to the Focus Unit*

The undeflected spot, even after careful setting of the lenses, showed an aberration resembling spherical aberration in which the marginal rays were over-focused when the paraxial rays were properly focused.[4] The effect was to give quite large symmetrical fan-shaped 'tails' in the horizontal plane on either side of the spot. This error was cured by shunting the peripheral field in the horizontal plane of the second lens, where the beam width was greatest, and hence most easily influenced (see Fig. 5(a)). This was achieved by sticking two vertical strips of iron at the front of the ring.

5.4. *Disadvantages of Magnetic Focusing*

5.4.1. Sensitivity of picture focus to e.h.t. fluctuations

The intermediate (virtual) line images produced by the prefocusing lenses are both situated within the scan magnifier field (Figs. 5(a) and (b)). Hence the effect of a small change in their axial position is to produce a large change in the position of the final image, and thus very marked defocusing of the picture on the screen. Due to the higher magnification produced by the scan magnifier in the horizontal plane, defocusing is much more apparent in this plane. For example, calculations show that a movement of the vertical image by 0·25 mm causes a change of final image position of about 100 cm. However, the small beam angle at the screen has to be taken into account in calculating the depth of focus of the system.

Since the strength of any magnetic lens is dependent on the beam voltage, this small change in the position of the intermediate images can be produced by a small fluctuation in the e.h.t. supply to the tube. The figures quoted in the previous paragraph for image movements would be produced by about a 1% variation of the e.h.t. Experience has shown that in practice with these systems, a deterioration of picture focus is noticeable for a change of e.h.t. of only 50 volts, or about 0·25%. Complicated stabilization circuits are therefore necessary to compensate for this effect.

5.4.2. The setting-up procedure

The fact that the magnetic focus unit requires a considerable number of adjustments, the setting-up of which involves a rather lengthy and complicated procedure, makes it an unattractive commercial

proposition. Also the physical bulk of the magnets is a nuisance when considering how they may be mounted. Although units have been built using smaller, lighter magnets, with the elimination of one or more of the adjustments involved in the unit described above (e.g. by having a fixed ring separation), the necessity for a more attractive method of focusing remains. Such an improvement can be obtained by going to electrostatic focusing.

6. An Electrostatic Quadrupole Focusing Assembly

6.1. *The Advantages of Electrostatic Focusing*

There are two great advantages to be gained by using electrostatic rather than magnetic focusing in scan magnifying tubes of this type. The first is that defocusing effects due to changes of e.h.t. can be allowed for by making the focus potential of the lenses track the e.h.t. If the ratio of focus potential to beam potential remains substantially constant, then the prefocusing lens strengths are stabilized. The virtual line images produced by these lenses therefore remain fixed in position, and the focus at the screen does not change over a wide range of e.h.t. values.

The second advantage is that the heavy and bulky magnetic components used for focusing as described above are no longer necessary. The requirement for complicated mounting and adjustment facilities is therefore relaxed. This makes tube replacement easier and initial setting up is less of a problem.

Two disadvantages which somewhat offset these advantages are the following. The reduction in the number of adjustments over a magnetically focused tube must be reflected in a requirement for more accurate manufacture and location of components within the tube. Any small inaccuracies in gun alignment, for example, will be magnified.

Secondly, the useful aperture of the electrostatic lenses must be less than that for the magnetic lenses. Hence it is likely that aberrations will be more of a problem. A smaller neck diameter, which could lead to a shorter tube, is not therefore desirable. For example, since some correction for aberration 'tails' was necessary in the second lens of the magnetic focus unit, it is almost certain that corrections will also have to be made to the second lens of an electrostatic focusing assembly, unless steps are taken to limit the beam width in this lens.

In the next section design details are given of a type of electrostatic focusing, which has successfully overcome the disadvantages of the magnetic unit.

6.2. *The Slot Quadrupole Lens*

The nearest practical approach to an ideal quadrupole field may be obtained using two pairs of oppositely-charged, hyperbolically-shaped plates (Fig. 11). This structure is awkward to make and rather difficult to mount accurately in a cathode-ray tube. A comparatively small proportion of the cross-sectional area of the neck is taken up by the useful aperture of the lens. Other lens designs, such as those using four curved plates with a circular instead of hyperbolic cross-section, or four rods, give good quadrupole fields near the axis of the tube while making fabrication easier.

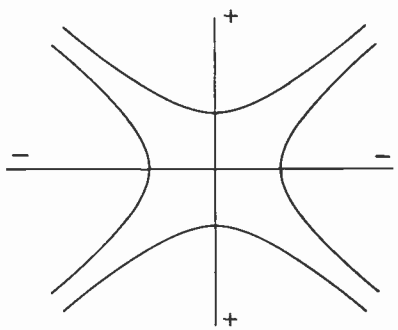


Fig. 11. Ideal electrode structure using hyperbolic plates for production of quadrupole field.

For an electrostatically focused scan magnifying television tube, a quadrupole design was chosen which offered the possibility of easy manufacture, and optimum use of the neck cross-sectional area. This lens is called a slot quadrupole. Basically it consists of two coaxial cylinders at different potentials, the inner of which has two similarly shaped and symmetrically disposed windows, or slots, cut into it. Figure 12 shows a diagram of the lens.

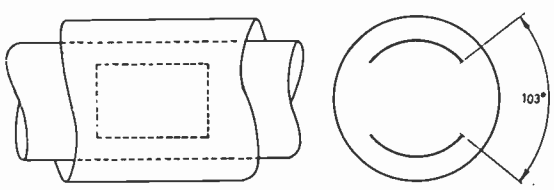


Fig. 12. Construction of slot quadrupole lens.

The required crossed quadrupole fields for the focusing system of a scan magnifying tube are then obtained by cutting a second pair of slots in the inner cylinder orientated at right angles to the first pair and at a suitable distance further along the tube. The two pairs of slots may use common or separated outer cylinders.

A pure quadrupole field (i.e. one excluding higher-pole components) extends only around and for a small distance away from the axis of the lens. Equipotential plots have shown that for lenses which have a ratio of inner to outer cylinder diameter approaching 1 the

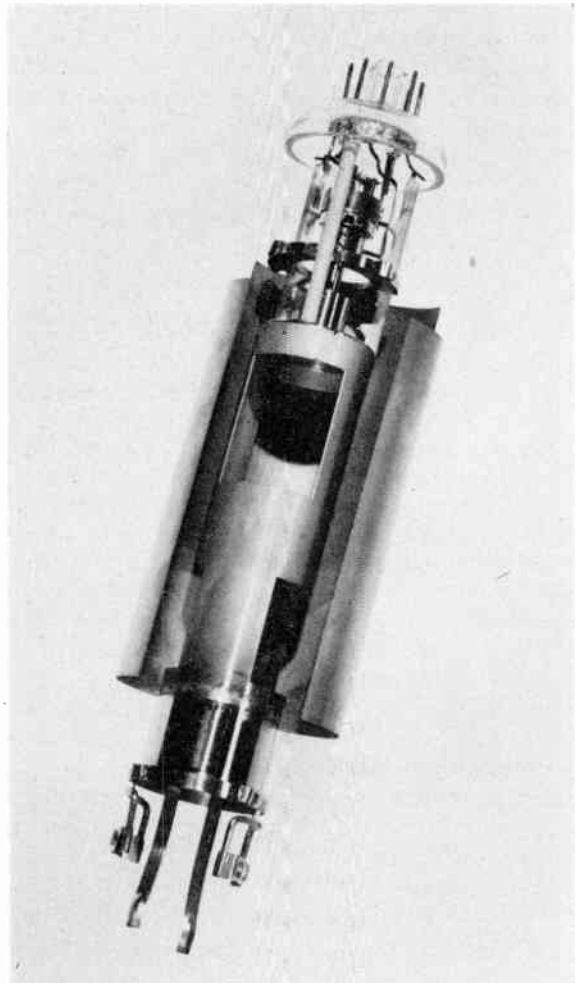


Fig. 13. Electrostatic focusing structure.

quadrupole field can extend with negligible error over an area defined by a circle of radius 0·4 times the inner cylinder radius. Outside this region higher-pole components (8-pole, 12-pole, etc.) begin to have a rapidly increasing effect on the field shape. The effect of these can be altered by varying the slot angle shown in Fig. 12 as 103 deg. This value was again determined from the equipotential plots for the condition of minimum deviation from a correct field over a maximum cross-sectional area around the axis.

With the slot angle fixed, the strength of the lens is determined by the axial length of the pair of slots, the diameter ratio of the cylinders and the potential difference between them. It is convenient to make the inner cylinder an extension of the electron gun anode so that it carries a high potential with respect to the earthed cathode; and to give the outer cylinder a potential near earth. Thus the separation of the cylinders can be fixed at a safe value to prevent spark discharges between them. For a given neck diameter, the maximum diameter ratio of the cylinders is therefore determined. This leaves only the slot length as the main variable in setting the correct lens strengths.

The lengths and positions of the slots were determined in the main by empirical methods. Some help was obtained from approximate calculations, while the results of ray-tracing experiments carried out at the Cambridge University Engineering Laboratory proved very useful.[5] However, the main method of approach was that of trial and error. A series of tubes was built, each one deriving from the previous one in the series until the correct lengths and positions of the slot pairs for a properly focused spot were found.

6.3. *The Construction of the Focusing Assembly*

The construction of a typical gun and focusing assembly is shown two-thirds of full-size in Fig. 13. The outer cylinder was formed from springy stainless steel foil 0·001 in thick so that it fitted snugly against the inside of the neck when the gun was inserted into the tube. It was fixed at the cathode end only, to a ring which was jigged into position with the other components and fixed to the ceramic insulators. This ring was used to provide a connection to the outer cylinder via a spare pin in the seven-pin base of the tube. It also located one end of the assembly in the neck. The inner cylinder was a stainless steel tube of 24·6 mm outer diameter, and 23·8 mm inner diameter. It was supported at the screen end by a stepped extension piece.

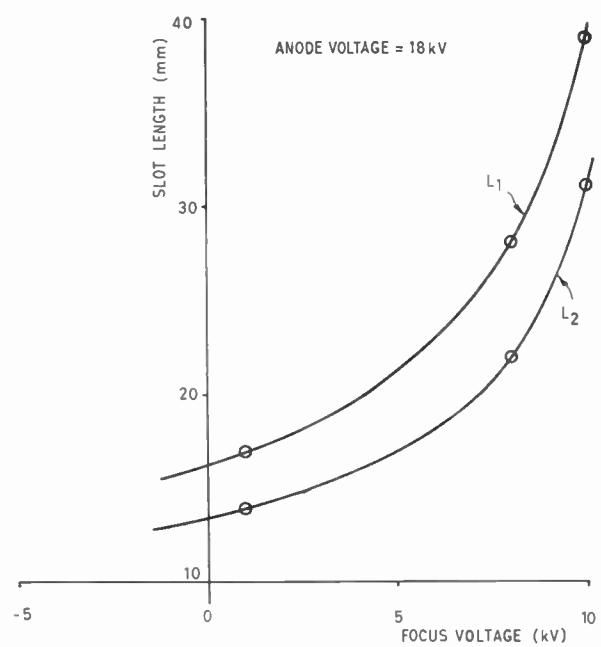


Fig. 14. Relation between slot length and focus voltage for electrostatic tubes.

Correct focusing was found to occur with the types of scan magnifier previously described when the grid plane of the gun was fixed at 280 mm from the screen, and the centres of the slots were at 48 mm and 87 mm from the grid plane. Figure 14 shows the slot lengths for three tubes with different focusing potentials. Trend curves of slot length L for the first and second quadrupoles against focusing potential have been drawn through the points. These curves correspond approximately to the rule that for a fixed lens strength

$$\Delta V \,.\, L = \text{constant}$$

where ΔV is the potential difference between the cylinders.

6.4. *The Need for a Second Focus Control*

In practice it was found that the single control of focus allowed by varying the potential on the outer cylinder common to both prefocus lenses was inadequate. Tubes of nominally similar dimensions had slightly different focusing properties due to small inaccuracies of construction. Experiments were carried out on tubes having separate outer cylinders for each prefocus lens, so that individual control of the lenses was possible. These tubes were, however, not easy to make as the outer cylinder for the second prefocus lens had to be separately supported in the tube, and an additional connection had to be taken through the foot of the tube.

A much simpler solution of this problem was to use a thin (1 mm) 'Magnadur' ring (52 mm outer diameter, 38 mm inner diameter), magnetized as a quadrupole and placed over either of the prefocus lenses to weaken or strengthen the lens as necessary to obtain correct focusing. The rotational position of this control had to be carefully adjusted for optimum results, but the axial position was not critical. There was a wide range of settings of its position and of the focus potential for satisfactory focusing. This control can be seen in Fig. 15, which shows a general view of an electrostatically focused tube together with the smaller scan magnifier.

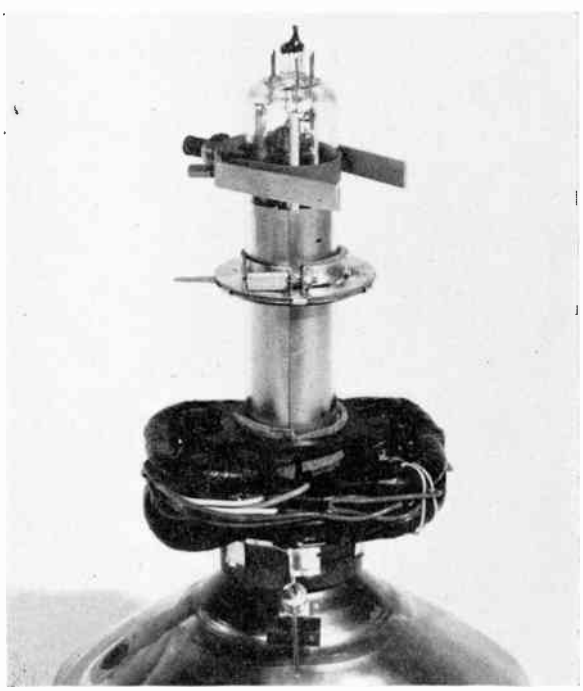


Fig. 15. Electrostatically focused tube.

6.5. *Correction of Aberrations in the Undeflected Spot*

When a focused spot was obtained at low beam currents as described above, aberrations were still observed at high beam currents. These were very similar in form to the fan-shaped tails produced by the magnetic focus unit. An attempt was made to provide correction by influencing the field in the outer regions of the second prefocus quadrupole, as with the magnetic unit. A series of tubes was built which incorporated a modified pair of slots for the second lens. These modifications took the form of wires fixed across the slots and subsidiary narrow slots cut into the inner cylinder between the main slots. These investigations, although showing limited success, were not carried to a final conclusion since the reduced field penetration through the slots caused by the modifications necessitated a much greater slot length. Some overlapping of the scanning fields with the focusing field was therefore inevitable at a point where the beam was already very wide. This meant that deflection defocusing problems were very much increased.

It was decided instead of this approach to insert a beam trimming aperture of 1·2 mm diameter into the anode before the first lens at 25 mm from the grid plane. The aperture was designed to limit the beam width in the second lens to about 7 mm or 0·3 times the inner cylinder diameter. Subsequent experiments showed that an estimated 50% of the beam current was lost due to trimming by the aperture at peak drive. The peak brightness of the tube therefore corresponded to about 500 μA beam current, or a dissipation of 9 watts at the screen. To allow for inaccuracies of gun construction, a beam positioning magnet placed near the grid plane was necessary so that the beam could be centred on the aperture.

7. The Deflector Coils and Deflection Defocusing

This Section describes the essential difference between deflector coils for conventional tubes and those for quadrupole scan magnifying tubes.

The main consideration, apart from sensitivity, in the design of deflector coils for a scan magnifying tube is the correction of deflection defocusing. In a conventional tube a compromise has to be accepted between this and raster distortion, the deflector coils

being designed to minimize both of these effects so far as possible. In a scan magnifying tube the corrections for deflection defocusing and raster distortion can be made largely independently of one another. As has already been explained, the correction of raster distortion is most conveniently carried out in the scan magnifier itself, as the coils have little effect on the raster shape compared with the lens. They do, however, have a considerable effect on spot shape and size with deflection.

These tubes were found to give negligible defocusing with horizontal deflection since the maximum angle of deflection necessary for complete scanning allowing for the scan magnification was very small (about 2 deg on either side of the axis). Thus only a small fraction of the scan magnifier aperture in this direction was used. The design of the line coils was therefore not very critical from this point of view. However, field coils designed to give a uniform deflecting field produced a marked horizontal defocusing of the spot with vertical deflection, i.e. the spots at top and bottom edges of the screen were horizontal ellipses in shape. If the beam voltage was increased, the spots at the edge of the screen became focused while the centre spots became defocused, showing that the coil and lens together were overfocusing the deflected beam. The coils were therefore redesigned to increase the defocusing action horizontally with increasing vertical deflection. The curvature of the image plane could then be made practically equal to that of the screen, so that good overall focusing was obtained.

The coils designed for the magnetically focused tubes applied vertical defocusing correction by means of two specially shaped 'Ferroxcube' cores separated by a spacer. The cross-sections of these cores are shown in Fig. 16. The angle between the field coils and the vertical plane was increased in the second ring nearer to the screen compared with the same angle in the first ring, the resulting change of field shape giving the required correction.

The deflector coils used with the electrostatic tubes were shorter than those for the magnetic tubes and a different method of defocusing correction was used. In this case the core consisted of two identical 'Ferroxcube' twelve-slot yokes stuck together without a spacer. The coils were positioned in the slots as shown in Fig. 17. Two subsidiary coils, marked C in the diagram, and wired in series with the field coils produced a distortion of field shape with increasing field current which corrected the defocusing at top and bottom of the picture. The reason for the shorter coils in this case was to avoid as far as possible deflecting over the latter part of the focus assembly, and introducing additional aberrations.

With the electrostatic tubes it was necessary to make some provision for centring the beam in the deflector coils and scan magnifier system. This was provided by picture shift magnets clamped just prior to the coils. This was largely unnecessary with the magnetically focused tube as it was possible to line up focusing and deflection systems individually.

A point of interest concerning the coils for scan magnifying tubes is that more 'S' correction of the line coils drive current appears to be necessary than for a conventional 90-deg tube. This is attributed to stronger edge fields with the scan magnifiers used here than would be the case if these lenses were perfect quadrupoles. The magnification factor thus increases rather than remaining constant with increasing deflection.

8. The Performance of Electrostatically Focused Tubes

In order to allow an estimate of the performance and capabilities of the electrostatically focused scan magnifying tube, this Section records some measurements and observations on a typical experimental tube.

The characteristics of cathode current and estimated screen current against grid drive voltage are shown in Fig. 18. The variation of line width with grid drive is given in Fig. 19. The line width was measured using a

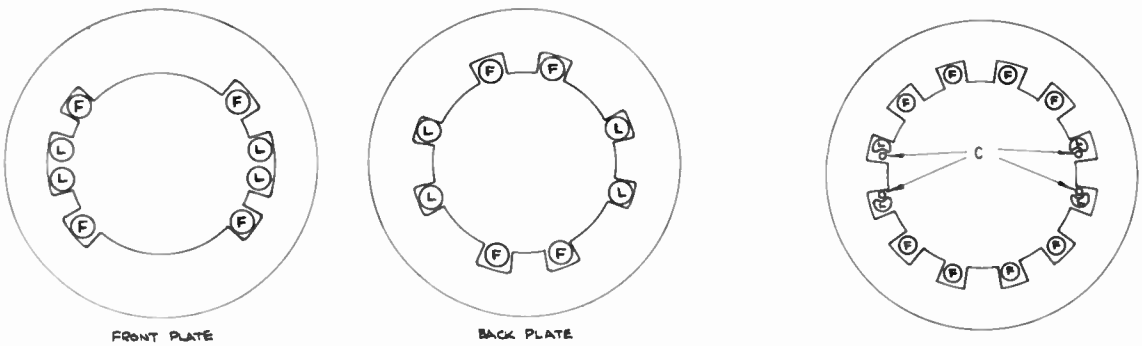


Fig. 16. Cross-sections through 'Ferroxcube' cores of one type of deflector coils.

Fig. 17. Cross-section through 'Ferroxcube' core of second type of deflector coils.

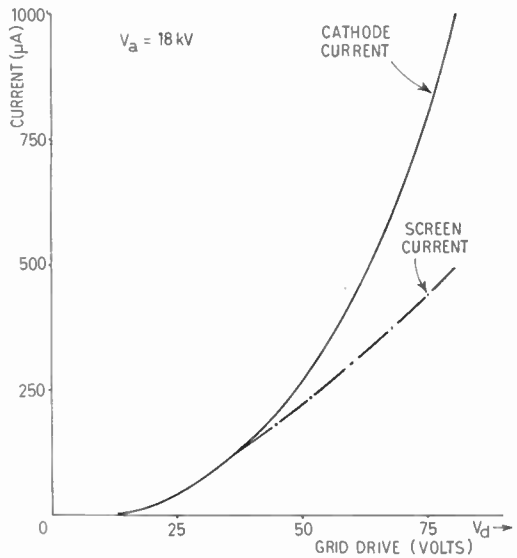


Fig. 18. Cathode current and estimated screen current against grid drive voltage for electrostatic tube.

travelling microscope to obtain the distance between points of visual extinction of light on a single line which just filled the centre of the screen in the horizontal direction, the focus controls having been adjusted for a round focused spot.

No provision was made in these tubes for separate screen and anode connections, so that the actual screen current for a given drive voltage could not be measured directly. The line plotted in Fig. 18 is an estimated screen current with grid drive characteristic based on measurements of the drive voltage at which the unfocused beam was just trimmed by the limiting aperture, and making the simplifying but somewhat doubtful assumption that the current density across the beam prior to trimming followed a Gaussian law.

A photograph of Test Card C received from a B.B.C. transmission (Channel 1) with a high quality r.f. receiver and displayed on the tube is shown in Fig. 20. The mean anode current for this picture was about 40 μA. This gave a measured mid-grey brightness of 32 foot-lamberts and a peak white brightness of 160 foot-lamberts.

The resolution, as indicated by the test bars in this picture did not appear to be markedly inferior to that of a conventional $8\frac{1}{2}$-in tube under the same conditions. This is better than would be expected from the line width measurement of Fig. 19. The key factor linking resolution and visual line width is the current distribution in the spot. Even at low beam currents, when the spot is circular in shape, it is believed that the distribution for this gun does not follow a strictly Gaussian law, especially towards the edges of the beam. At high beam currents, the effects of aberration 'tails' in the horizontal plane are apparent in the picture, which shows a certain loss of 'crispness'.

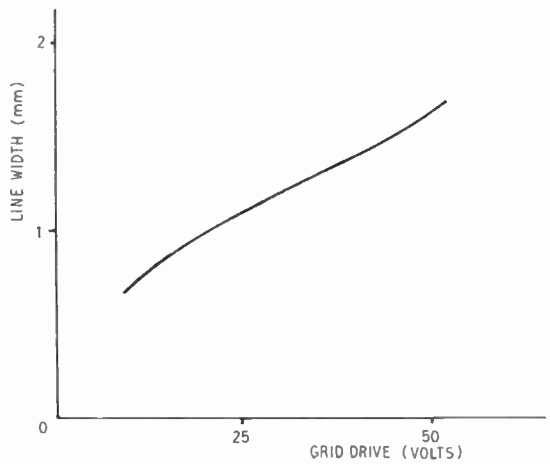


Fig. 19. Visual line width against grid drive voltage for electrostatic tube.

The picture shows the difficulty of correcting the raster exactly; although the edges are straight, other parts of the picture towards the middle are incorrect by small amounts. These errors are inconspicuous in most picture material.

This tube used the smaller and lighter scan magnifier and required only a thirtieth of the line deflection power of a conventional 18 kV 90-deg tube, while needing about the same field deflection power as the conventional tube. A magnetically-focused tube using the heavier scan magnifier and longer deflector coils required only a hundredth of the conventional line power and about a half the field power. Although further development of the electrostatic tube could undoubtedly improve the above quoted figures for this type, those for the magnetic tube give an approximate upper limit for the electrostatic tube, since the deflector coil length is limited by the need to avoid deflecting over the second focus lens.

In contrast to the magnetic tubes, the e.h.t. for the electrostatic tubes could be supplied from an unstabilized generator. Tests showed that it was

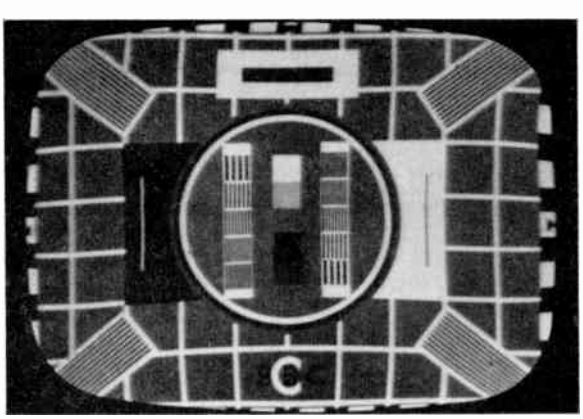


Fig. 20. 405-line Test Card 'C' displayed on electrostatic tube.

possible to vary the screen potential by about 3 kV on either side of the nominal 18 kV without serious defocusing of the picture, providing the focus, or outer sleeve, potential tracked the e.h.t. This range adequately copes with the typical output impedance (3 to 6 MΩ) of a conventional e.h.t. generator supplying about 1 mA peak anode current.

An important point in connection with picture definition was noted during tests of electrostatic tubes with different slot lengths. Tubes with shorter slot lengths produced poorer spot quality and picture definition than those with longer slots. This was attributed to increased aberrations due to the increasing influence on total lens properties of the edges of the slots, perpendicular to the tube axis, as slot length is decreased. The approximation to the required quadrupole field becomes increasingly worse with decreasing slot length with consequent deterioration of spot quality. This effect has an unfortunate practical repercussion in that the preferable longer slots require a smaller potential difference between anode and outer sleeve. Hence a high focus potential is associated with the longer slots which can lead to difficulties concerning maximum safe voltage rating when the lead-out from the outer sleeve passes through the foot of the tube.

9. Conclusions

Scan magnifying tubes using quadrupole lenses have been shown to require less than a thirtieth, and possibly as little as a hundredth, of the line scanning power of conventional tubes having the same deflection angle and anode voltage. Field scanning power can also be reduced to about half of that normally required.

These scanning power reductions may be achieved using permanent magnetic scan magnifiers made from sintered 'Magnadur' focusing rings. A simple method of magnetizing these rings has been given. The rings are mounted on the tube neck between the deflector coils and the bulb, the same mounting serving both components. The only adjustment necessary is provision for a small rotation of the rings relative to the coils so that the diverging plane of the ring can be made coincident with the line deflection plane.

The scanning coils are of simple design and present no greater difficulty in manufacture than conventional coils. Two different techniques of deflection defocusing correction have been described in connection with the coils, both techniques providing satisfactory correction in practice.

Raster correction of the display may be provided partly by the method of magnetization of the scan magnifier rings and partly by means of external small bar magnets placed on or near the bulb of the tube. The adjustment of these small magnets has been found to be critical for a satisfactory picture. This error has proved consistently to be the most difficult one of the display to correct, and it is not claimed that the solution of this problem given in this paper provides more than an acceptable compromise.

Two types of focusing, a permanent magnetic form and an electrostatic form, have been described in detail. The permanent magnetic form, consisting of two 'Magnadur' rings like the scan magnifier lenses and magnetized in a similar way, gives satisfactory focusing but suffers from two severe disadvantages: it is extremely difficult to adjust and requires a much more stable e.h.t. supply than is normally provided in conventional television receivers.

The electrostatic type of focusing to a very large extent overcomes these difficulties. Providing the focus voltage tracks the anode supply, no more stabilization of e.h.t. is required than in a normal receiver. The extra cost of the potentiometer chain to enable this tracking to be achieved, plus the existence of a high voltage on the 'foot' of the tube and the consequent need for insulation of the focus control are disadvantages introduced by the method. Adjustment of a beam positioning magnet over the grid plane and a small focus correcting ring are necessary. Each of these presents no greater difficulty than the adjustment of a conventional 'ion trap' magnet.

Because the scan magnifier magnifies the effects of constructional errors as well as the deflection, much smaller tolerance limits on gun alignment and centrality are necessary with these tubes than with conventional tubes. The experimental tubes were made to very tight limits and little information has been gained on the effect of larger errors in manufacture. However, experience has indicated that while the gun and electrostatic lenses were comparatively cheap and easy to make, the failure rate due to badly off-centre or badly aligned guns was high. In production the limits on alignment might have to be as much as an order tighter than for conventional tubes.

The picture quality of the experimental tubes left something to be desired, particularly with reference to resolution and geometry. However, laboratory experience has shown that tubes using electrostatic focusing can produce acceptable quality on most picture material.

An assessment of the total power requirements of a 625-line receiver using an electrostatic tube of the type described in this paper has been made, the figure arrived at being 8·7 watts. A similar conventional tube (AW21–11), operated at 12 kV, produced a figure of 17 watts. One of the main reasons for the scan magnifying receiver not producing a greater power saving than this 2 : 1 reduction is the necessity for the provision of a separate e.h.t. and video h.t. supply generator. This fact, coupled with the increased cabinet length necessary because the scan magnifying

tube is 40% longer than the AW21–11 (305 mm compared with 215 mm), and the fact that it is more complex and less easy to manufacture and adjust than conventional tubes, has led the author to the conclusion that quadrupole scan magnification has little to offer at the present stage of development of television tubes.

10. Acknowledgments

The author wishes to thank the Directors of Mullard Ltd., for permission to publish this paper. He also wishes to express his deep indebtedness to Mrs. M. L. Beeforth, Messrs. D. R. Skoyles, R. H. C. Morgan, and M. G. Hulyer. Thanks are also due to Dr. K. F. Sander and Mr. B. C. Gregory of the Cambridge University Engineering Laboratory for their help in measuring the focusing properties of slot quadrupoles.

11. References

1. N. W. Parker, I. A. Caorba and H. N. Frihart, "Recent developments in scan magnification", *Inst. Radio Engrs Int. Conv. Rec.*, **8**, No. 7, p. 167, 1960.
2. P. J. Dolon and W. F. Niklas, "Electron transmission of mesh lenses for scan magnification in television picture tubes", *J. Brit.I.R.E.*, **20**, p. 911, December 1960.
3. B. R. Overton, "Transistors in television receivers", *J. Television Soc.*, **8**, No. 11, 1958.
4. J. C. Burfoot, "Third order aberrations of doubly symmetric systems", *Proc. Phys. Soc.*, **B67**, p. 523, 1954.
5. B. C. Gregory and K. F. Sander, "A study of some properties of an electrostatic quadrupole lens using an automatic electron trajectory tracer", *J. Electronics Control*, **13**, No. 2, pp. 123–36, August 1962.

12. Appendix: Conditions for Magnification of Scan in the Vertical Plane

Choosing x and y axes to lie in the mirror planes of symmetry of the field shown in Fig. 1, the plane $y = 0$ being the diverging plane, and writing $\eta = \sqrt{(e/2m)}$ where e/m is the ratio of electronic charge to mass, the equations of motion of electrons moving at small angles to the tube axis are:

$$\sqrt{\phi}x'' = -\eta y' H_z + \eta H_y;\ \sqrt{\phi}y'' = \eta x' H_z - \eta H_x \quad \text{.....(1)}$$

where $'$ indicates differentiation with respect to z, and ϕ is the constant electrostatic potential in the region considered, i.e. no electrostatic field exists. In the vertical plane, $x = 0$, of a quadrupole field, the field components in the co-ordinate directions are $H_x = a_{11}\,y$, $H_y = H_z = 0$; and in the horizontal plane, $y = 0$, $H_y = a_{11}x$, $H_x = H_z = 0$, where a_{11} is a function of z only. So the equations of motion become

$$\sqrt{\phi}x'' = \eta a_{11}x;\ \sqrt{\phi}y'' = -\eta a_{11}y \quad \text{.....(2)}$$

In general, since a_{11} is a function of z these equations cannot be solved simply. However, if a_{11} is assumed to be a rectangular function of z, that is, it is constant in a region defined by a length L (approximately the physical length of the lens) and is zero outside this region, then the equations have simple solutions. If the origin of z is taken to be the boundary of the lens which the electron meets first, and the position and direction of the electron at this boundary are given by (x_0, y_0) and (x'_0, y'_0) respectively, the solutions of equations (2) are

$$\left.\begin{aligned} x &= (x'_0/\lambda)\sinh \lambda z + x_0 \cosh \lambda z \\ y &= (y'_0/\lambda)\sin \lambda z + y_0 \cos \lambda z \end{aligned}\right\} \quad \text{.....(3)}$$

where

$$\lambda^2 = \eta a_{11}/\sqrt{\phi} = \sqrt{e/2m\phi}\,(\partial H_x/\partial y) = \sqrt{e/2m\phi}\,(\partial H_y/\partial x)$$

λ is called the quadrupole constant.

If the rays make angles x'_L, y'_L with the axis as they leave the lens (in the plane $z = L$) the angular magnifications of scan in the two co-ordinate planes are given by

$$\left.\begin{aligned} \frac{x'_L}{x'_0} &= \cosh \lambda L + \lambda \frac{x_0}{x'_0}\sinh \lambda L \\ \frac{y'_L}{y'_0} &= \cos \lambda L - \lambda \frac{y_0}{y'_0}\sin \lambda L \end{aligned}\right\} \quad \text{.....(4)}$$

If the deflection centres for both planes occur at the same point on the axis, say at $z = -d$,

$$x_0/x'_0 = y_0/y'_0 = d$$

and

$$\left.\begin{aligned} x'_L/x'_0 &= \cosh \lambda L + \lambda d \sinh \lambda L \\ y'_L/y'_0 &= \cos \lambda L - \lambda d \sin \lambda L \end{aligned}\right\} \quad \text{.....(5)}$$

The angular magnification in the vertical plane is obviously an oscillating function with a first maximum value occurring when $\pi/2 < \lambda L < \pi$. The stationary values of this function are given by the solutions of the general equation

$$\frac{\tan \lambda L}{\lambda L} = -\frac{d}{L+d} \quad \text{.....(6)}$$

If d and L both have the approximate practical value of 20 mm, then the first solution of (6) gives a value of 2·3 for λL. The corresponding maximum value for the vertical angular magnification is $-2{\cdot}4$ times, the negative sign indicating that the beam has become inverted. The horizontal angular magnification is then 16 times.

A similar analysis for the linear magnification observed on a flat screen placed 10 cm from the end of the scan magnifier gives values of $-1{\cdot}8$ times for the vertical plane and 13 times for the horizontal plane.

Manuscript received by the Institution on 15th February 1964. (Paper No. 920/T28.)

U.D.C. 621.396.677

Theoretical and Experimental Properties of Two-element, Multiplicative Multi-frequency Receiving Arrays including Superdirectivity

By

B. S. McCARTNEY,
Ph.D. †

Reprinted from the Proceedings of the Symposium on "Signal Processing in Radar and Sonar Directional Systems", held in Birmingham on 6th–9th July, 1964.

Summary: The paper presents the theory and properties of the directional responses of two-element multiplicative arrays, sometimes termed correlation arrays, when receiving multi-frequency signals. The results of an experimental investigation employing band-pass signals are given and agree well with the predicted directional responses. Interesting modifications to the straightforward multi-frequency arrays are demonstrated, including a split-beam response, arrays with one channel clipped, and superdirective arrays. Superdirectivity is examined in terms of the improved effective aperture, for which a quantitative definition is proposed. Some degree of multi-frequency superdirectivity was obtained, but further gain of effective aperture is restricted by the accuracy to which the amplitudes of the received frequencies may be controlled.

1. Introduction

There has been considerable interest of late in wide-band receiving arrays with various geometries and processing methods. For sonar systems the possibility of obtaining good directivity with arrays of small dimensions by employing wide bandwidths is particularly attractive and worthy of investigation. Some results of an investigation into two-element, multiplicative arrays, sometimes called correlation arrays, are presented here and compared with theory.

1.1. *The Space-frequency Analogy*

The idea of the analogy between the frequency complexity of a wide-band, two element, multiplicative array and the spatial complexity of a linear, additive array to give identical directional patterns was first put forward by Kock and Stone[1] in 1957. In another paper Kock[2] gave some examples (Fig. 1), and demonstrated a practical multi-frequency array using sound in air. Welsby[3] also gave an example using a multi-frequency transmission and suggested[4] a means of scanning such an array electronically. Tucker[5] demonstrated that the analogy is not exact in several respects, but that superdirectivity could be applied. In that paper it was also shown that a multi-frequency, two-element additive array could be produced but with considerable complexity.

Consider a two-element correlation receiving array as shown in Fig. 2, receiving a plane wave from a direction θ, with a pressure waveform $p(t)$. Assuming ideal, omnidirectional transducers the output voltages $v_1(t)$ and $v_2(t)$ from elements 1 and 2 will be proportional to $p(t)$ and $p(t+\tau)$, where

$$\tau = \frac{2d \sin\theta}{c} \qquad \text{.....(1)}$$

The integrator output may be written as

$$D_{11}(\tau) = \frac{1}{2T}\int_{-T}^{+T} v_1(t).v_1(t+\tau)\,dt \qquad \text{.....(2)}$$

where T is the response time or averaging time of the filter or integrator. Since τ is a function of θ, $D_{11}(\tau)$ is the far-field directional response of the array. The auto-correlation function $\psi_{11}(\tau)$ of the waveform $v_1(t)$ is defined by

$$\psi_{11}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} v_1(t).v_1(t+\tau)\,dt \qquad \text{.....(3)}$$

It may be seen that there is a very close resemblance between $\psi_{11}(\tau)$ and $D_{11}(\tau)$, so much so that in many cases they can, with caution, be treated as proportional.

The Wiener-Khintchine theorem states that the auto-correlation function $\psi_{11}(\tau)$ is the Fourier transform of the power density spectrum $E_1(\omega)$ of the waveform $v_1(t)$, so that, by identifying $D_{11}(\tau)$ with $\psi_{11}(\tau)$, it follows that the directional function of this array and the power density spectrum of the received signal are Fourier transform pairs. It is well known that for a plane aperture at a single frequency, the sensitivity function $S(y)$, relating

† National Institute of Oceanography, Wormley, near Godalming, Surrey; formerly with the Department of Electrical Engineering, University of Birmingham.

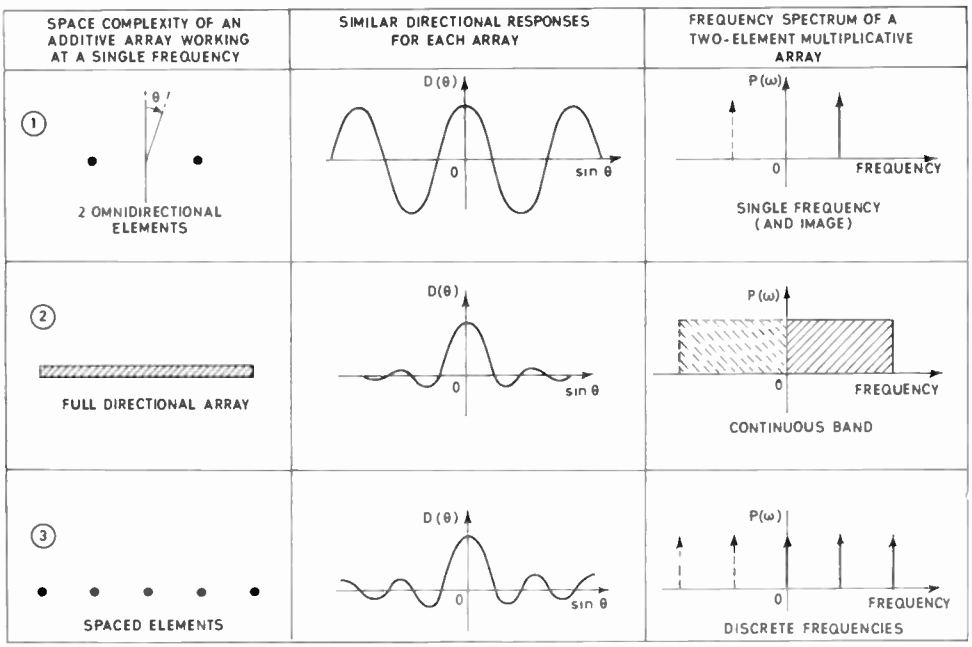


Fig. 1. Examples of the space-frequency equivalence (after W. E. Kock[1]).

sensitivity to position along the aperture, and the far-field directional pattern $D(p)$ or $D(x)$, are Fourier transform pairs. Hence, by direct analogy, the directional patterns $D(x)$ and $D_{11}(\tau)$ may be related, providing also the sensitivity function $S(y)$ is related to the power density spectrum $E_1(\omega)$. For example, a linear point array receiving a single frequency is analogous to a two-element, multiplicative array receiving a multi-frequency signal.

The physical limit to the aperture sensitivity $S(y)$ is analogous to a bandwidth limitation on $E(\omega)$, which is a realistic physical limitation also. However, whereas the aperture limit of a plane array is an abrupt one, that on $E_1(\omega)$ must, in general, be a gradual one. In the space-frequency analogy, y and ω are the related variables; y may be positive or negative since position is measured relative to the centre of the aperture but ω usually has meaning only for positive values. By introducing the concept of negative frequencies this may be overcome. However, the power density spectrum $E_1(\omega)$ is inherently a positive even function of ω, whilst $S(y)$ may be negative, non-symmetrical or even complex. This restriction on $E_1(\omega)$ amounts to restricting $D_{11}(\tau)$ to an even function with a maximum value at $\tau = 0$.

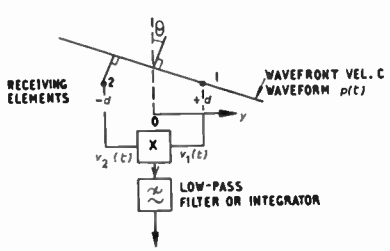


Fig. 2. Simple auto-correlation array.

It should be noted that $p(t)$ and $v_1(t)$ need not be precisely proportional and in practice will not be so due to the complex transfer function of the transducers. If the elements are identical and each channel contains identical networks, then $E_1(\omega)$ and $E_2(\omega)$ will be identical and

$$E_1(\omega) = P(\omega)|H_1(j\omega)|^2 \qquad \text{.....(4)}$$

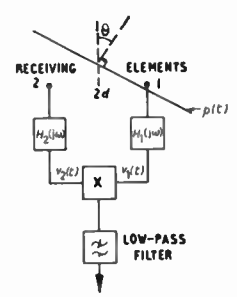


Fig. 3. Simple cross-correlation array.

where $P(\omega)$ is the power spectrum of $p(t)$ and $H_1(j\omega)$ is the transfer function of a complete channel, transducer and network together. Having chosen a

positive even function for $E_1(\omega)$, analogous to a required sensitivity function $S(y)$, it may be achieved by controlling either $P(\omega)$, $H_1(j\omega)$ or both; that is we may choose the waveform $p(t)$ or network to derive the required taper and hence directional response.

By inserting different networks into the signal paths before multiplication (Fig. 3), $v_2(t) \neq v_1(t+\tau)$ and the multiplier output becomes,

$$D_{12}(\tau) = \frac{1}{2T}\int_{-T}^{+T} v_1(t).v_2(t+\tau)\,dt \qquad .....(5)$$

which now resembles the cross-correlation function;

$$\psi_{12}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} v_1(t).v_2(t+\tau)\,dt \qquad .....(6)$$

The Wiener-Khintchine theorem states that the cross-correlation function $\psi_{12}(\tau)$ and the cross-power density spectrum $E_{12}(\omega)$ are Fourier transform pairs. The cross-power density spectrum does not have the same physical significance as the power spectrum for the auto-correlation case, but can still be usefully related to the analogous sensitivity function $S(y)$. It may be shown that

$$E_{12}(\omega) = P(\omega).H_1^*(j\omega).H_2(j\omega) \qquad .....(7)$$

where $H_1^*(j\omega)$ denotes the conjugate complex function of $H_1(j\omega)$. The number of equivalent taper functions which may be applied is increased now to include positive and negative amplitudes and tapers with odd symmetry. Thus the peak response may now be deflected from $\tau = 0$ and split-beam and superdirective responses may be obtained. Again either $P(\omega)$, $H_1(j\omega)$ or $H_2(j\omega)$ may be used to control $E_{12}(\omega)$, and the phase responses of the networks become important. It may not always be possible to design a realizable network to give the required taper.

The space-frequency analogy enables the theory of the directional responses of plane apertures to be applied to the multiplicative correlation array, with some limitations and differences discussed in Section 1.2. In particular the analogy enables the well-known theory of linear point arrays to be applied to the multi-frequency, two-element array. The waveform $p(t)$ must satisfy the wave equation, but otherwise there are an infinite number of functions which have the same $P(\omega)$ because the phase spectrum is unimportant; this fact enables the designer to have some control over the ratio of r.m.s.-to-peak value of $p(t)$, which may be important when high power transmitters are employed near the limits of cavitation or electrical breakdown. The simplest form of $p(t)$ is periodic, in which case the limits of the integrals of equations (3) and (6) may be made finite and equal to the half period of $p(t)$. There is exact agreement then between $\psi_{11}(\tau)$ and $D_{11}(\tau)$ and similarly between $\psi_{12}(\tau)$ and $D_{12}(\tau)$ assuming that the integration is perfect. In this (multi-frequency) case the integration amounts to filtration of the zero frequency products from all other frequencies, the nearest of which is the periodic frequency or its second harmonic. In the cases of aperiodic or random, band-limited noise-like waveforms of $p(t)$ the ideal infinite integration time cannot be achieved in practice and the integration time chosen is a compromise between the output noise bandwidth and the pulse length if range information is required.

1.2. *Some Properties of Multi-frequency Arrays*

Only the properties of correlation arrays which are not identical to those of linear additive arrays are mentioned here.

(1) Whereas $D(\tau)$ appears as a direct voltage which may be positive or negative, $D(x)$ arises from the amplitude of an alternating voltage, and after detection, is positive only. Synchronous demodulation is rarely of practical significance, but would allow $D(x)$ to have either polarity like $D(\tau)$.

(2) The multi-frequency array operating over a wide bandwidth is susceptible to frequency dependent transmission irregularities which cause distortion of the directional pattern. The echo from a moving target has a Doppler frequency compression or expansion. These frequency shifts produce a symmetrical change in the pattern equivalent to a change in angular scale or element spacing. The effect will usually be small since target speeds are low compared with the speed of sound. However, if band-pass operation as described in Section 1.3.1 is used, the Doppler shift is approximately the same for all frequencies and so the harmonic relationship is destroyed; also the shift is a larger proportion of the frequency spacing.

(3) The directional response $D(\tau)$ is also distorted symmetrically by any frequency dependence of the target echo strength. Prior knowledge of this effect may facilitate target recognition, but would more likely cause confusion.

(4) The correlation array directional output is not formed until after the multiplier which is an irreversible device, so that it is clear that reciprocity in the normal sense does not apply. A multi-frequency transmission does not produce a distribution of power in space according to the directional response on reception.

(5) The multiplication process introduces several secondary effects which further prevent accurate analogy; the output level is proportional to the square of the input signal level, though this could be overcome if necessary by a non-linear amplifier; more important

is the fact that the response of the array to anything other than a single source is complex and cannot be obtained by simple superposition, because of cross multiplication terms. The responses with multiple targets, multipath propagation and noise are mentioned in Section 5.

1.3. *Some Practical Aspects of Multi-frequency Arrays*

1.3.1. Band-pass working

If the power spectrum $E_1(\omega)$ is finite in a pass-band, which does not extend down to zero frequency, and is zero outside the band, the directional function $D_{11}(\tau)$ is, by analogy, the same as that of an additive interferometer of two linear arrays, and has rapid alternations of polarity within a broader beam. When the signal phase is non-stationary, rectification could be applied after multiplication and smoothing but the usual advantage with multiplicative arrays of easily removing the negative side-lobes is lost.

A more practical alternative for working within a pass-band in the medium is to frequency-shift the band on reception, prior to multiplication. In this case the signal transmitted is not $p(t)$, but a carrier modulated by $p(t)$. The type of modulation is not important as far as the directional function is concerned, but under some conditions advantages may be obtained using one form in preference to another. Frequency modulation generally requires a wider transmission bandwidth than amplitude modulation for the same modulation bandwidth, but could lead to improved performance with regard to multiple-path effects and some forms of noise by the use of strong clipping. With amplitude modulation, multi-path signals may cause partial or complete cancellation of the carrier and cause the detected output to be a very distorted version of $p(t)$ plus the multi-path signal. The probability of this distortion occurring is reduced the smaller the modulation index, in which case most of the transmitted energy is in the carrier and not the sidebands.

Band-pass working is the most practical form for experimental multi-frequency transmission and reception, as the need for wide-band transducers, or many narrow-band ones, is reduced to one narrow-band transducer per channel. The practical investigation described below used narrow-band working with amplitude modulation.

1.3.2. Further modulation of $p(t)$

An active sonar requires range as well as bearing information and modulation of $p(t)$ would be required for range determination. It is clear that the bandwidth, $\delta\omega$ of any modulation of $p(t)$ must be restricted and that the multiplier doubles this bandwidth at the output. Due to this inherent modulation distortion by the multiplier, the use of a multi-frequency directional system in a communication link is limited in general to the additive array described by Tucker[5] (shown in Fig. 4 of this reference), in which the simplicity of the correlation array is lost. If the information modulation on $p(t)$ is in some form of pulsed code, the distortion may not be serious. However, besides distortion of the signal modulation, the presence of this modulation causes a distortion of the response; this distortion is reduced as $\delta\omega$ is made much smaller than the fundamental frequency ω_1, or in a pulsed system as the pulse length T_p is made much greater than the period T_1.

When signals arrive on zero bearing the pulses in both channels coincide, so that the output pulse has the same length; ideally the pulse length need only be equal to T_1 the period of the fundamental, in order to allow accurate integration. However, from other directions, the pulses do not coincide, so that the pulse output of the multiplier is shorter, having a minimum length of $(T_p - 2d/c)$. There will thus be many directions for which the averaging process is made over a non-integral number of periods of the fundamental so distorting the directional response. Without specifying the degree of distortion which may be tolerated, an exact limit cannot be placed upon $\delta\omega$, which, in any case, must be limited to avoid overlap of the distorted spectra, according to:

$\delta\omega < \omega_1/4$ when odd and even harmonics are used

or

$\delta\omega < \omega_1/2$ when odd harmonics only are used.

It is clear from the restrictions of $\delta\omega$ mentioned above that full use is not made of the complete overall transmission bandwidth because of the gaps in the spectrum. It may be shown[6] that the ratio $\delta\omega/\omega_1$ is further restricted when certain forms of electronic scanning are applied. To minimize the input receiver noise and medium noise, a comb filter accepting only the narrow bands around each harmonic would be needed, but this would rarely be necessary in a reverberation-limited sonar.

1.3.3. Directional responses of individual elements

The directional response of a multi-element, single frequency array is the product of the array space factor and the response of an element individually, all elements being assumed equal. Having calculated the directional response of the correlator pair assuming omnidirectional elements over the full operating band, account must be taken of any directionality of the two elements.

When the two transducers have equal, or unequal but constant, directional responses with frequency, the overall array output is the product of their responses and the cross-correlation function. However, when the element responses are frequency sensitive,

$H_1(j\omega)$ and $H_2(j\omega)$ become functions of angle of arrival and so the equivalent array taper varies with the direction of arrival. This would be particularly troublesome with superdirective multi-frequency arrays (see Section 4). When band-pass working is used for transmission and reception, the directional responses of normal transducers would vary little over the band.

2. Experiments and Results

2.1. *Experimental Array and Apparatus*

When Kock and Stone[1] verified the idea of a multi-frequency receiving array in air, standard audio microphones and loudspeakers were used which had wide bandwidths. Though wide-band, low-frequency hydrophones are available for underwater reception, the problem of obtaining a wide-band transducer which will radiate controlled acoustic power into water, is difficult and has not been satisfactorily solved yet. Accordingly, for the convenience of transmission, band-pass operation described above was chosen for the experimental work.

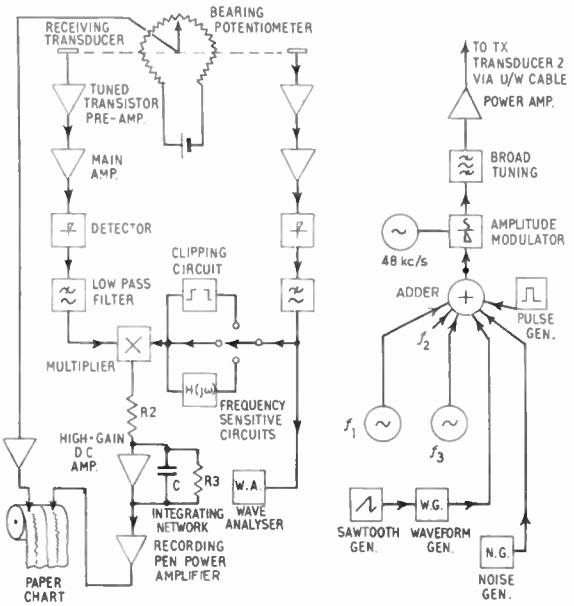


Fig. 4. Block diagram for experimental work.

Some magnetostrictive transducers resonating at 48 kc/s were available having a Q of approximately ten, giving a 3 dB bandwidth of 4·8 kc/s. In order to obtain a reasonable directional response which could be compared with theory a minimum of three odd harmonics was required. Thus the fundamental modulation frequency was 500 c/s. Each receiving element consisted of a rectangular stack having measured, 3 dB beamwidths of 44 deg and 10 deg in perpendicular planes; they were fixed with their long axes perpendicular to a horizontal transducer bar, the spacing of the elements being variable from about 2 inches to 10 feet ($0{\cdot}02\lambda_1$ to λ_1, where λ_1 is the wavelength of the fundamental modulation frequency). The transducer bar could be rotated through 360 deg and a potentiometer facilitated measurement of the array bearing. The undermentioned experiments were performed in a reservoir, and a wooden jetty provided a working platform for the receiving array. The appropriate acoustic signal was transmitted from a transducer out in the reservoir towards the receiving array fixed to the jetty. The ideal medium in which to measure directional array responses would be infinite or else anechoic and of course the reservoir is neither of these. Considerable time was spent in obtaining an arrangement which was as free from multiple paths as possible. At the fundamental modulation frequency of 500 c/s the water would be considered shallow, though at the carrier frequency of 48 kc/s much less so. Since the major reflection interference had such a small time difference from the direct signal, it was concluded that nothing could be gained by using pulse transmissions for measuring the directional responses and c.w. was used throughout.

Block diagrams of the receiver and transmitter are given in Fig. 4. The majority of the circuit design was straightforward and requires little comment. For the multiplier a ring multiplier[7] with crystal diodes was used. This rather simple circuit uses low input levels to a balanced diode ring so that the semiconductor diodes operate over the portion of their charac-

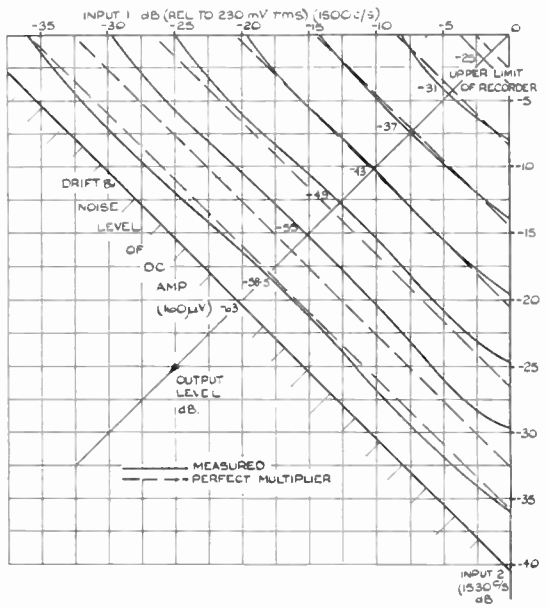


Fig. 5. Ideal and measured characteristics of diode ring multiplier.

teristic which is exponential. Four-quadrant operation is obtained and the zero d.c. output level is stable.

The minimum usable output level is determined by the effective noise or drift level at the input to the high gain d.c. amplifier, which follows the multiplier, and the maximum by the level at which the diodes' characteristics depart from the exponential law. Cathode followers fed transformers to provide the necessary balanced drive to the ring. The characteristics of the multiplier are plotted in Fig. 5, where one input level is plotted against the other to give a constant output level. It can be seen that, over a small dynamic range of each input, the multiplier is reasonably accurate, but at extremes non-linearities appear. Since in this application the input levels are approximately the same, and it is possible to restrict the input levels to a limited range, this moderate performance was accepted. The d.c. amplifier used was a Solartron AA 621·2, high gain, computer amplifier with drift compensation and the resistor-capacitor network (R2, R3, C) conveniently performed the required integration. The ratio $R3/R2$ determined the d.c. gain and $R3 \times C$ determined the integration time.

An Evershed and Vignoles paper recorder was used, having two pens fed by d.c. power amplifiers with overall feedback. Their response times were each approximately 100 ms. One channel recorded the array output and the other channel simultaneously recorded the array bearing from a voltage on the wiper of the potentiometer. The bearing calibration was 5 deg per division (2 mm). Both channels became slightly non-linear for large deflections of the pen. This was partly counteracted by non-linear paper (Fig. 6) and all measurements were made within the linear range.

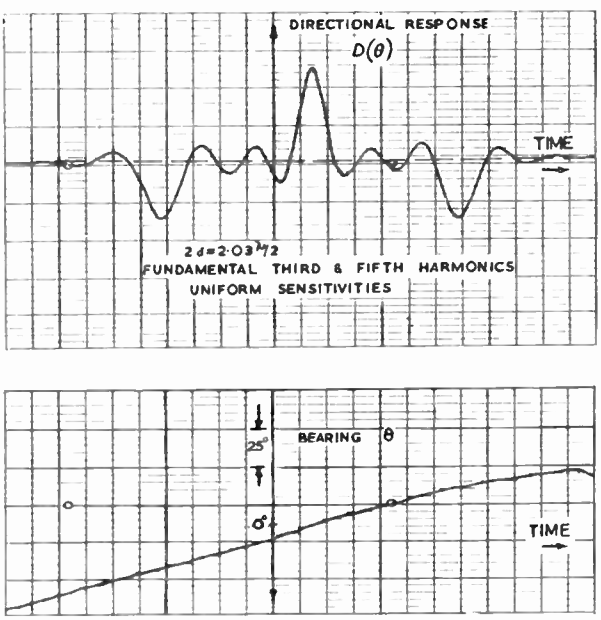


Fig. 6. Example of chart recording.

2.2. *Measurement Errors*

The maximum departures from linearity of the bearing potentiometer were ±3 deg, as measured on a resistance bridge and, owing to the thickness of the ink trace, there is a reading error estimated as ±1 deg, giving a maximum bearing error of ±4 deg.

The main sources of amplitude error in the receiver are the detector, the multiplier and any differential phase-shift between channels, which is not a simple time delay. In the transmitter, non-linearity of the modulator and errors in setting the array harmonic taper functions are the main causes of error.

A 48 kc/s carrier, amplitude modulated with a 1000 c/s tone to 50% by the transmitter modulator, was injected into a receiver input transformer and the detected output measured (Fig. 7). It is seen that a good linearity is obtained for a range of 25 dB,

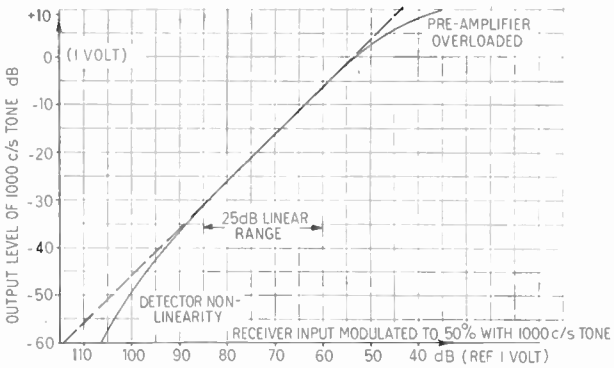


Fig. 7. Receiver detector linearity.

which is adequate for the present purpose. From Fig. 5 it may be seen that the multiplier error varies from −1 dB at high levels to +2 dB at low levels over most of the area produced by the top 20 dB of each input signal. Each point on the graph of the overall frequency response (Fig. 8) represents a long term average reading in the presence of some fluctuations of the order of ±5%. Since the amplitudes of the harmonics determine the array taper function, in conjunction with the square of the frequency response, any error in the latter will be approximately doubled when applied to the taper function. Though this error could be ±10% for each harmonic, it is likely that the actual errors were much less than this figure. The array spacing between centres was measured to $\pm\frac{1}{8}$ in which was better than ±1% at the lowest spacing used. The velocity of sound in water was assumed to be 4750 ft/s.

The differential phase-shift between the receiver channels was not discernible with an injected signal

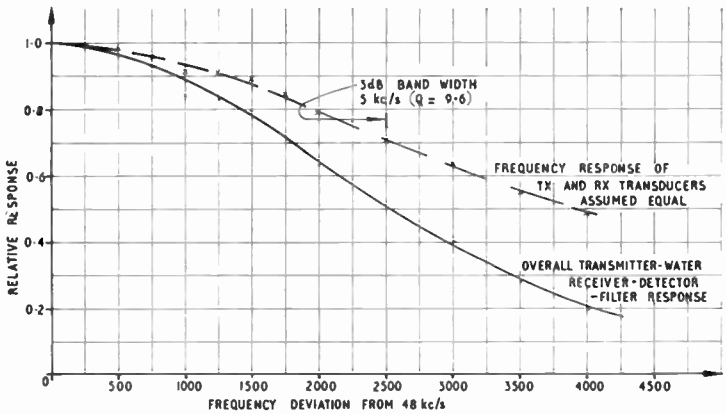


Fig. 8. Measured frequency responses.

but further phase-shifts are introduced by the transducers and could cause errors. Also a non-symmetrical phase response about the carrier frequency would introduce phase and amplitude distortion of the detected output.

2.3. *Form of Results*

Having decided upon the array to be tested, the amplitudes of the modulation harmonics were set accordingly, noting that the harmonic voltage levels needed to be proportional to the square root of the desired taper coefficient. The overall modulation index was then set to 50%. The array was slowly rotated manually and the directional response and bearing were recorded simultaneously (Fig. 6). This trace was the best and is free from medium fluctuations, but many others were not quite so fortunate. Where direct comparisons with theory have been made (Figs. 9 and 10), the directional responses have been re-plotted on a linear scale of bearing θ, normalized in level, and some of the modified theoretical points inserted (circled points); the modified theoretical points are taken from the theoretical response multiplied by the product of individual element responses. Since the latter product falls to 10% of its

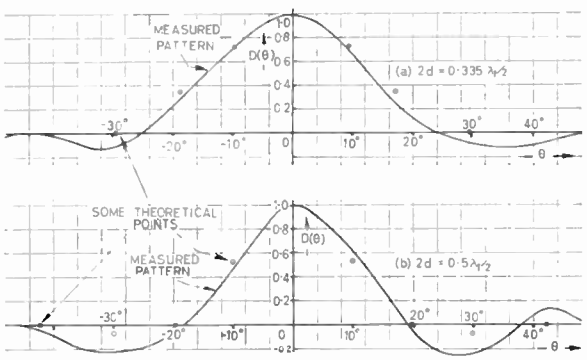


Fig. 9. Measured multi-frequency directional patterns (1st, 3rd, 5th harmonics with array spacing indicated).

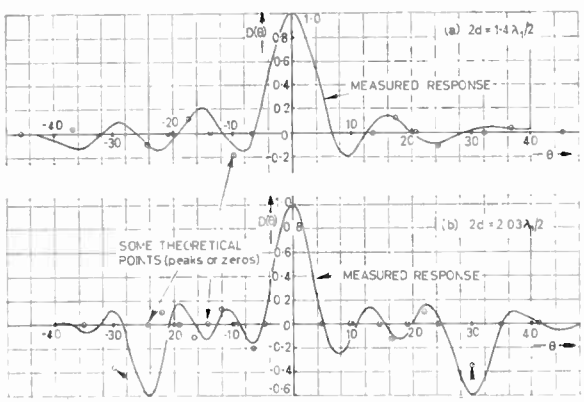


Fig. 10. Measured multi-frequency directional responses (1st, 3rd, 5th harmonics).

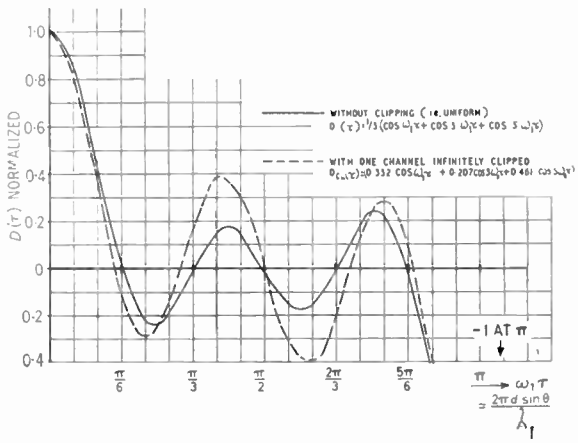


Fig. 11. Directional patterns for 3-frequency array with and without clipping in one channel.

peak value at ±40 deg, this determined the sector of measurement. Where the trace merely illustrates an effect, only the directional response trace is given, with the bearing indicated every 5 deg (Figs. 12 to 16). The various arrays attempted and the patterns obtained are given below.

2.4. *Straightforward Multi-frequency Arrays*

In order to confirm the multi-frequency idea with band-pass working, uniform arrays with fundamental, third and fifth harmonics were chosen initially and the array spacing was varied from $0{\cdot}335(\lambda_1/2)$ to $2{\cdot}03$ $(\lambda_1/2)$. The basic directional response of such an array is of course $\frac{\sin 6p}{\sin p}$, where $p = \omega_1\tau = 2\pi (2d/\lambda_1) \sin\theta$ and which is shown in Fig. 11. A six-element, single frequency, linear array has the same form but $p = \pi(d_0/\lambda_0) \sin\theta$, where λ_0 is the wavelength of the applied signal, and usually the element spacing, d_0, is quoted in multiples of λ_0; thus to align the angular scales of the two arrays requires that the multi-frequency array spacing, $2d$, is the same multiple of $\lambda_1/2$ as the multi-element array spacing is a multiple of λ_0. Figures 9 and 10 show the results for spacings of $0{\cdot}335(\lambda_1/2)$, $0{\cdot}5(\lambda_1/2)$, $1{\cdot}4$ $(\lambda_1/2)$ and $2{\cdot}03(\lambda_1/2)$. As the element spacing is increased, the beamwidth narrows until eventually secondary principal maxima appear as expected. As the individual element responses do not have a zero, the theoretical and measured zeros should coincide. In fact most of the measured zeros do occur within a few degrees of the theoretical points, all the exceptions being at large bearing angles. As the true speed of sound was not known, any error in its value would give a slightly incorrect value for $2d(\lambda_1/2)$ and owing to the $\sin\theta$ dependence of p this would be more apparent as a bearing error at larger values of θ. There is a tendency for the first side-lobe level to be lower on the negative side than on the positive side, which suggests the presence of a small phase error. The maximum amplitude errors occur at high bearings and are positive. Owing to the reduced signal level into the multiplier channels at these bearings due to the reduced transducer responses, a positive error is expected which may be as large as 3 dB relative to the peak response. Also the errors due to fluctuations in the measurements of transducer responses and overall frequency response are particularly noticeable at low levels and therefore the higher bearings. In view of these facts, the patterns of Figs. 9 and 10 are considered to be within the bounds of experimental error and it is concluded that there is no serious practical deviation from the theory of multi-frequency directional responses using band-pass working. The arrays described in Sections 2.5, 2.6 and 2.7, further confirm the idea, and were obtained mainly for interest.

The patterns discussed above were obtained using co-phasal, phase-locked harmonics on transmission, but no difference was observed when phase-locked harmonics which were not co-phasal were used; nor indeed even when frequencies from separate oscillators were used providing the latter were set to within a few c/s of the required frequency.

2.5. *Odd and Even Harmonics*

The transmission of the second and fourth harmonics in addition to the three odd harmonics produced the response of Fig. 12, the taper function

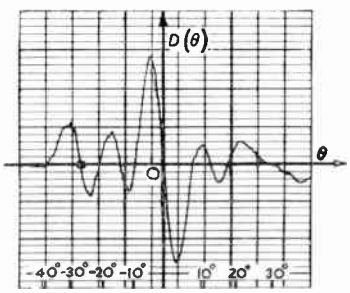


Fig. 13. 'Split-beam' response.

(c_r^2) being unintentionally non-uniform as indicated. However, the response is interesting in that it has very few and small positive side-lobes and is correct for the taper. A slight modification to the taper would give a response without positive side-lobes. Had the array been uniform its response would have been $\left\{\frac{\sin 11\omega_1\tau/2}{\sin \omega_1\tau/2}\right\} - 1$, equivalent to an eleven-element uniform array with the centre element missing.

2.6. *Split-beam Response*

For accurate bearing measurement, it is sometimes useful to use a 'split-beam' directional response, which has a null and high slope at zero bearing. A split-beam, multi-frequency array was formed by the

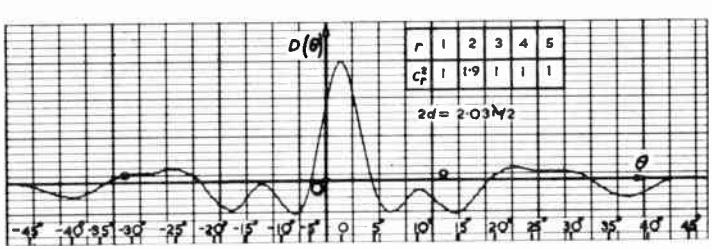
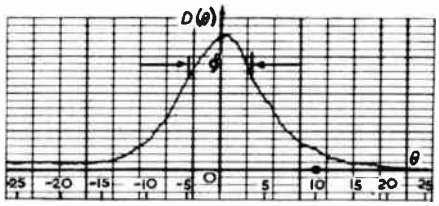
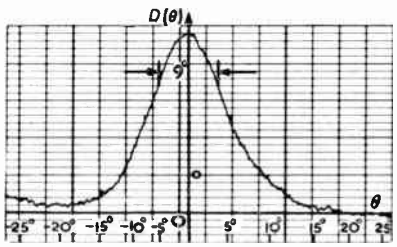


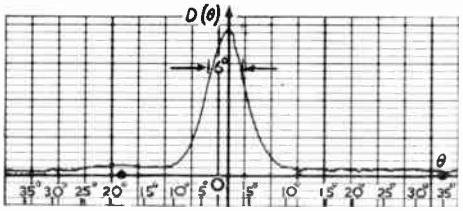
Fig. 12. Odd and even harmonics (non-uniform taper).



(*a*) $2d = 5'$



(*b*) $2d = 5'$ (Channel B clipped)



(*c*) $2d = 10'$

Fig. 14. Directional patterns with band-limited noise.

insertion in one channel of circuits giving 90 deg phase shift at each harmonic frequency. The response of Fig. 13 was obtained. The amplitude taper was not uniform, but if it had been uniform, the response would have been

$$\psi_{12}(\tau) = \sum_{r=1,3,5} \sin \omega_r \tau = \left(\frac{1-\cos 6\omega_1 \tau}{\sin \omega_1 \tau}\right)$$

which clearly has odd symmetry.

2.7. *Band-limited-noise Arrays*

As is clear from the space-frequency analogy the use of noise for the transmitted modulation is analogous to a continuous array. The unfiltered output of a white noise generator was used to modulate the carrier, so that the equivalent array sensitivity function was the square of the overall frequency response (Fig. 8). A good analytical approximation to the squared curve is the Gaussian curve given by,

$$E_1(\omega) = \frac{1}{\sqrt{2\pi}.\omega_\sigma} \exp\left(\frac{\omega^2}{2\omega_\sigma^2}\right)$$

whose Fourier transform, and therefore the directional response in this case, is given by the similar curve,

$$\psi_{11}(\tau) = \exp(-\omega_\sigma^2 \tau^2/2)$$

ω_σ is the 'standard deviation' of frequency and is $2\pi \times 1{\cdot}5 \times 10^3$ rad/s here. The response falls to $1/e = 0{\cdot}368$, when

$$\omega_\sigma \tau = \sqrt{2}$$

i.e. when

$$\theta_e = \sin^{-1}\left(\frac{\sqrt{2}c}{\omega_\sigma . 2d}\right)$$

$$= 8{\cdot}1^\circ \quad \text{when } 2d = 5{\cdot}08 \text{ ft}$$

and $\theta_e = 4{\cdot}1^\circ$ when $2d = 10$ ft

From the measured responses in Fig. 14 (*a*) and (*c*) it may be seen that there is good agreement at these points and the curves are of approximately Gaussian form. The complete absence of side-lobes, however wide the spacing, is an advantage, though of course a longer integration time is required than with a few harmonics.

3. Hybrid Clipped Arrays: Co-phasal and Non-cophasal

Some advantages may be obtained by performing the multiplication in digital rather than analogue form. The conversion from analogue to digital form implies some degree of amplitude quantization, as well as time-sampling at each input, which can introduce error into the correlation process. Recently, Watts[8] has shown that these errors may be quite small even for coarse quantization. The extreme degree of quantization for correlation is obtained when each input is 'infinitely' clipped, presented to a phase-sensitive bridge, or sign comparator, and then averaged.

The peak output level of a true analogue correlator is proportional to the product of the two input r.m.s. levels; this product is used to normalize the correlation function to form the correlation coefficient. Clipped correlation, though it is not true correlation, is automatically normalized to be independent of input signal level. In the absence of a signal, the output is low providing the cross-correlation of the noise and reverberation at the two receiving elements is small. One of the main advantages of clipped correlation is the ease[9] with which signal delay may be introduced using shift registers. The disadvantages of clipped correlation are the loss of signal level information and a small reduction in the output signal/noise ratio.

Some of the advantages of both digital and analogue correlation may be obtained by a hybrid method: one channel is clipped and then used to switch a modulator, to which the other, unclipped channel is applied as signal. The output after filtration has a peak correlated level proportional to the incoming signal level. The largest disadvantage with this hybrid

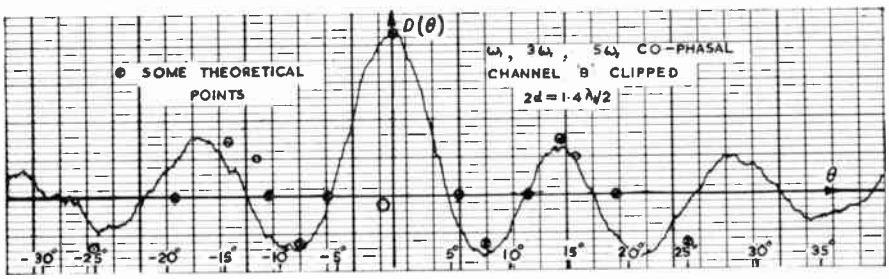


Fig. 15. Hybrid clipped array with co-phasal harmonics.

system is that the directional response is now dependent upon the relative phase of the signal harmonics. This is more significant when only a few harmonics are received than when a large number or a noise-like wave-form is used. The directional response for a simple case in which co-phasal fundamental third and fifth harmonics are received has been calculated and is shown in Fig. 11, together with the uniform unclipped multi-frequency array for comparison. It is seen that the maximum secondary lobe level of the clipped hybrid array is approximately twice that of the uniform unclipped array, though the beamwidth is not significantly reduced.

A co-phasal waveform was transmitted and the response obtained (Fig. 15) is quite close to that predicted by Fig. 11 multiplied by the response of one channel alone, and again a phase error is apparent. The experiment was repeated except that the phases were not co-phasal, though locked, and the result is given in Fig. 16. Figures 15 and 16 may be compared with Fig. 10(*a*) which is the unclipped response at the same array spacing.

This type of clipping can also be applied with a band-limited noise-signal, and the response in Fig. 14(*b*) is that obtained as for (*a*) except that one channel was clipped. Any reduction in beamwidth expected due to clipping is offset by the increase in beamwidth due to multiplication by only one transducer response, but again the response is free from side-lobes.

In general the effect of clipping on the spectrum of a signal is to increase the high frequency power content at the expense of the low frequency power content. When both channels are clipped this is analogous to increasing the aperture of a single-frequency multi-element array, at the expense of the sensitivity at the centre; this tends to result in slightly reduced beamwidths, increased side-lobe levels and poorer signal/noise gains, particularly at low input signal/noise ratios. In a hybrid array where only one channel is clipped the energy transferred to high frequencies in the clipped channel does not contribute to the output of the multiplier and so the effective array aperture is not increased, though the sensitivity distribution is altered.

4. Superdirectivity and Multi-frequency Arrays

Before giving the results of some multi-frequency arrays with superdirective tapers, some discussion of superdirectivity itself is necessary. The terminology of multi-element, single-frequency arrays is used.

It is possible to obtain, with a finite aperture, any far-field directional response at a fixed frequency, providing the sensitivity distribution may be accurately and continuously controlled over the aperture. In practice it may be extremely difficult to achieve such a distribution, but frequently an approximation may be made by dividing the aperture into n elements, whose individual sensitivities are controlled; in this

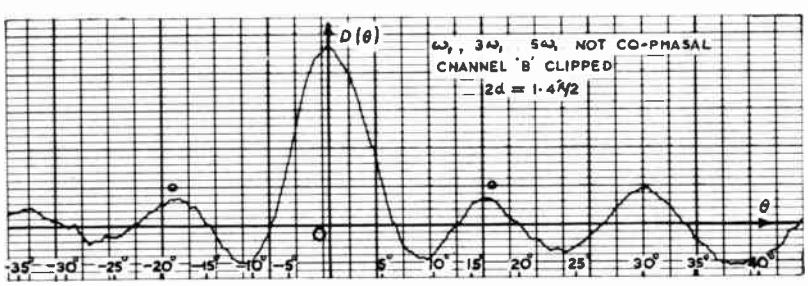


Fig. 16. Hybrid clipped array with non-co-phasal harmonics.

way the directional response is specified in n arbitrary directions. All practical methods of synthesis of linear arrays consist of calculating the array coefficients from the specification of the response in n directions. Advantages may be obtained by using unequal spacings, but for simplicity equal spacings are assumed here. As the element spacing becomes less than $\lambda/2$, the sensitivity distribution may be designed to produce what are known, by common usage, as superdirective or super-gain arrays. No formal definition of superdirective arrays exists, but they have some or all of the following properties:

(i) Their effective aperture is greater than their physical aperture.

(ii) Their 'directivity factor' (d.f.) is greater than that of a uniform array of the same length.

(iii) Their absolute sensitivity must be greater than that of a uniform array to give the same maximum output (i.e. their efficiency is low).

(iv) Their sensitivity distributions oscillate along the aperture with a spatial period of less than a wavelength, and alternate elements may require reversed phase connection. (In the case of n even the two centre elements have the same phase.)

(v) The bandwidth over which there is no appreciable change in directional response and maximum output is reduced.

(vi) They have large responses in the range of complex angles, that is those for which $|\sin\theta| > 1$.

(vii) They have a poor 'noise factor' (n.f.) as a result of (vi).

It is instructive to examine these properties in relation to each other and to the properties of non-superdirective arrays. An important property to a designer is the value of the d.f. and this would be a very useful basis for the design of an array. Except for the case when $D(\theta)$ is very small when $\theta \approx \sin\theta$, the change of variable in $D(x)$ or $D(p)$ to $D(\theta)$ before integration makes the computation of d.f. very difficult. In general, therefore, it is not convenient to approach the design of superdirective arrays from their d.f. The property (i) above is the broad definition of superdirective arrays used by Tucker[10]; this has the advantage of amounting to (ii) for arrays which are required to produce a narrow fan beam and may also be used for bearing determination arrays of the split beam type. A method of defining 'effective aperture' is now proposed, which may be used whatever the actual response and which facilitates quantitative comparisons of the degree of superdirectivity and examination of the above properties.

The 'effective aperture' of a superdirective (or any) array may be defined as 'the length of the smallest array having its n_1 elements spaced $\lambda/2$ apart, which will give the same response at n_1 directions within the range of real angles as the n element superdirective array'. No restriction has been placed on either the number of elements n_1, or their sensitivity distribution, so that in principle an equivalent array with half-wavelength spacing may be found for any superdirective array. For the present purpose, it is considered sufficient to specify the limited number, n, of directions at which the arrays shall correspond. The problem usually will be to obtain a short superdirective array having the same real angle response as a given longer $\lambda/2$-spaced array. Since the superdirective array has a directional pattern very similar to, if not exactly like, that of the reference array, their directivity factors must be similar, and the ratio of their lengths is an indication of the degree of superdirectivity applied. It is convenient to consider as parameter the ratio of superdirective element spacing, d_k, to $\lambda/2$ as a parameter designated $k = 2d_k/\lambda$). The aperture improvement ratio (I) may then be defined as n_1/kn. Two examples of superdirective

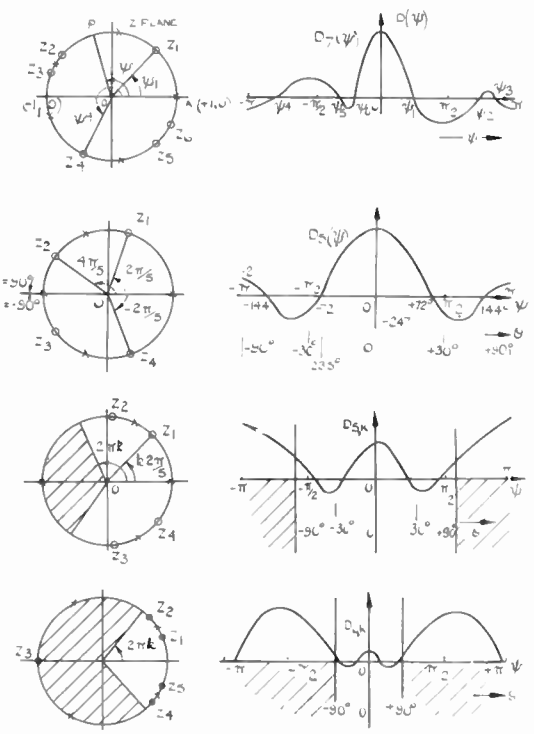


Fig. 17. Z-plane plots and corresponding $D(\psi)$ directional functions.

arrays having $n = 5$, or $n = 6$, have been calculated using Schelkunoff's method[11] of synthesis and using the response of a uniform $\lambda/2$-spaced array having $n_1 = 5$ as a reference.

4.1. *Examples of Superdirective Arrays*

The directional response $D_n(z)$ (more strictly $|D_n(z)|$) of an n-element array may be expressed as a polynomial of degree $(n-1)$ of the unit vector $Z(= e^{j\psi}$, where $\psi = 2p = \pi k \sin\theta$).

$$D_n(z) = a_0 + a_1 z + a_2 z^2 + \ldots a_{n-1} z^{n-1}$$

which may be written as the product of $(n-1)$ factors,

$$D_n(z) = a_{n-1}(z-z_1)(z-z_2)\ldots(z-z_{n-1})$$

Thus $D_n(z)$ has zeros at

$$z = z_1, z = z_2, \ldots z = z_{n-1}$$

corresponding to

$$\psi = \psi_1, \psi = \psi_2, \ldots \psi = \psi_{n-1}$$

and hence to

$$\theta = \theta_1, \theta = \theta_2, \ldots \theta = \theta_{n-1}$$

respectively. The positions of these zeros or nulls are indicated by the points $Z_1, Z_2, \ldots Z_{n-1}$ on the unit circle, Fig. 17(*a*). The range of real angles, θ, corresponds to $|\sin\theta| \leqslant 1$, i.e. $|\psi| \leqslant 2\pi d/\lambda$. Thus for an array with elements spaced at d_k the range of ψ is $\pm k\pi$ and of course for superdirective arrays $k < 1$. Clearly for a symmetrical response, the null points must be placed symmetrically relative to the real axis of the unit circle. Also for a principle maximum at $\theta = 0$, there must be a maximum at point A ($\psi = 0$), but all other points, P, may only be related to θ when k is known. The notation used considers the directional function to be expressed as $D_{n,k}(\theta)$ or $D_{n,k}(\psi)$, where n indicates the number of elements, and k the superdirectivity parameter previously defined. Figure 17(*b*) gives the zero and maxima distribution for the five-element uniform array with $k = 1$ and shows the corresponding $D_{5,1}(\psi)$ plot; there are four zeros in the complete range of 2π on the unit circle. In order that $D_{n,k}(\theta) = D_{5,1}(\theta)$ over the range $|\sin\theta| \leqslant 1$, the range of ψ from $-k\pi$ to $+k\pi$ must be divided in the same proportions as the whole circle for $k = 1$. However if n_1 had been even there would have been a zero at $\psi = \pi$ when $k = 1$, which would require two zeros, one each at $\pm k\pi$, when $k \neq 1$. The minimum value of n is therefore (n_1+1) when n_1 is even and n_1 when n_1 is odd. For this reason it is clear that an even number of elements would rarely be used in a superdirective array. For a split-beam type of response an odd number of zeros is needed in the range of θ, so that an even number of elements is then an advantage. From the point of view of the d.f. it is a waste to put a zero outside $\pm k\pi$ though it does of course improve the n.f. The z plane plots and corresponding $D_{n,k}(\psi)$ plots are shown in Fig. 17(*c*) and (*d*) for $n = 5$ and $n = 6$. The latter example, besides illustrating an even numbered superdirective array provides a convenient example for use in the multi-frequency experiment. Having specified the positions of the nulls $Z_1, Z_2 \ldots Z_{n-1}$, the $(n-1)$ array coefficients $(a_0\ a_1 \ldots a_{n-z})$ may be calculated[6] as functions of k and a_{n-l} will be chosen to normalize the array output to unity at $\theta = 0$. The coefficients for $n = 5$ are shown in Fig. 18 and in Fig. 19 for $n = 6$. It may be seen that all coefficients increase toward infinity more rapidly than I as k approaches zero. Also in each case a_1 and its symmetric element changes polarity as expected and then rapidly increases in the negative direction. The 5-element array becomes a 3-element array with double spacing when $k = 0{\cdot}833$, for which $a_1 = a_3 = 0$; similarly, the 6-element array becomes a 4-element array with unequal spacings in the ratio 2 : 1 : 2, when k = 0·685, for which $a_1 = a_4 = 0$; clearly neither of these designs would normally be termed superdirective.

A superdirective response arises from the small algebraic sum of large amplitudes, so that a small percentage change in the sensitivity of one or more elements has a much greater adverse effect than for a

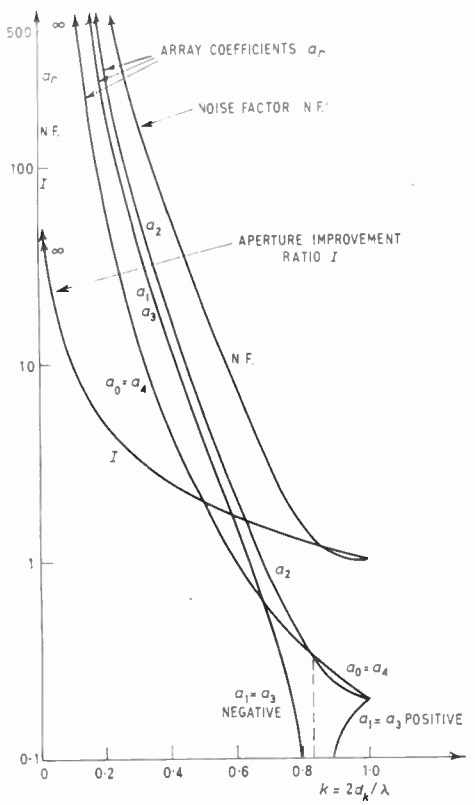


Fig. 18. Superdirective array functions.

$D_{5,k}(\theta) = D_{5,1}(\theta)$, $n = 5$

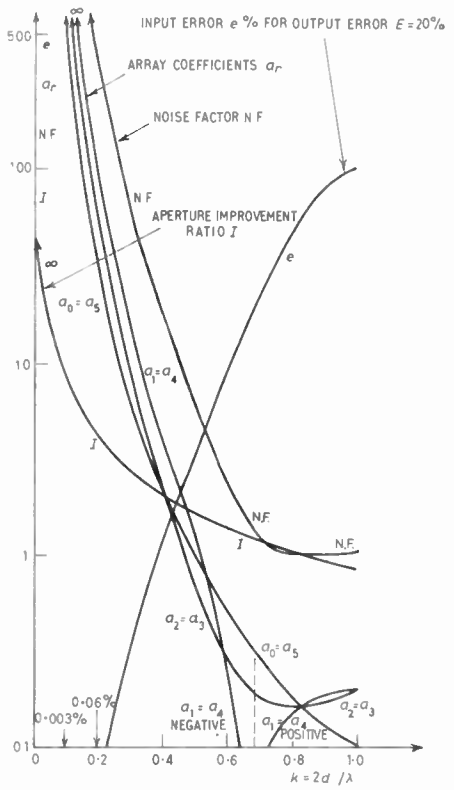


Fig. 19. Superdirective array functions.

$D_{6,k}(\theta) = D_{5,1}(\theta)$, $n = 6$

normal array. The modulus input change e% in each coefficient which gives at worst a change of $E = 20\%$ in the output at $\theta = 0$ is also shown in Fig. 19.

Since the slopes of the coefficient curves are high it is clear that the output response is likely to be severely affected by small changes in d_k and/or λ and thus in operating frequency or the velocity of sound. This is an indication of the reduced bandwidth over which a superdirective array may operate without appreciable change in output.

The noise factor is a measure of the performance of the array to noise which arises separately in each element and is uncorrelated; thermal noise is in this category and, unlike noise from the medium, does not have its phase-angles restricted to $k\pi$ between elements. Thus, due to the larger responses outside the range of real angles up to $|\psi| = \pi$ when $k < 1$, it is expected that the n.f. should deteriorate (i.e. increase). This is borne out by the curves of n.f. in Figs. 18 and 19, calculated from the array coefficients. Small but worthwhile aperture improvements of 20 to 30% may be obtained without difficulty for increases in n.f. of the order of 3 dB.

4.2. *Some Practical Difficulties with Superdirectivity*

So far the discussion has implicitly concerned receiving arrays; if passive linear components are used to produce the required taper, reciprocity is valid and the transmission response would be superdirective. The practical difficulties of achieving a superdirective array depends to a very large extent, upon whether receiving or transmitting, acoustic or electromagnetic arrays are involved.

The directivity and noise factors clearly have no meaning for a transmitting array, but the analytical presence of large responses outside the range of real angles still has practical significance; in this case, polarity reversals from one element to the next, spaced less than $\lambda/2$ away, produce a highly reactive near-field. This means that the impedances loading the driving sources are reactive; these would normally be tuned out, a further indication that narrow bandwidth is inherent in superdirectivity. Since losses in the elements will be proportional to the square of the circulating current, which is large compared with the component producing the far-field radiation, the power efficiency will be lowered as k is reduced. In water the peak displacement is limited by cavitation and so the power radiated into the far-field is limited to a smaller value than for a normal array. It has been assumed in the array analysis that all the elements are independent and do not interfere with each other or the incoming wave, but it is clear that as k is reduced coupling will occur between elements. The existence of mutual coupling is closely related to the reactive field.

4.3. *Multi-frequency Superdirectivity*

The extra requirements to make a multi-frequency array superdirective are a special filter[5] in one channel, a spacing $2d = k\lambda_1/4$ when odd harmonics are used, and of course the appropriate taper. The spacing of the elements at some of the multi-frequency harmonics may thus exceed half a wavelength and it is thought that some of the mutual coupling effects described may be avoided. Mutual interference amongst harmonics may now be caused by non-linear circuits including an imperfect multiplier. When band-pass working is employed mutual coupling is no longer a problem until k is extremely small as the carrier wavelength is much smaller than λ_1. The purpose of the special phase-shift network is to reverse the polarity of alternate harmonics. Several types of circuit may be used[6] and in the experiments an active circuit was chosen for its simplicity.

Three harmonics were again used, equivalent to a 6-element array, and the taper coefficients were obtained from Fig. 19. The results obtained, using band-pass working, are shown in Fig. 20, for the four values of k indicated. The theoretical response (T)

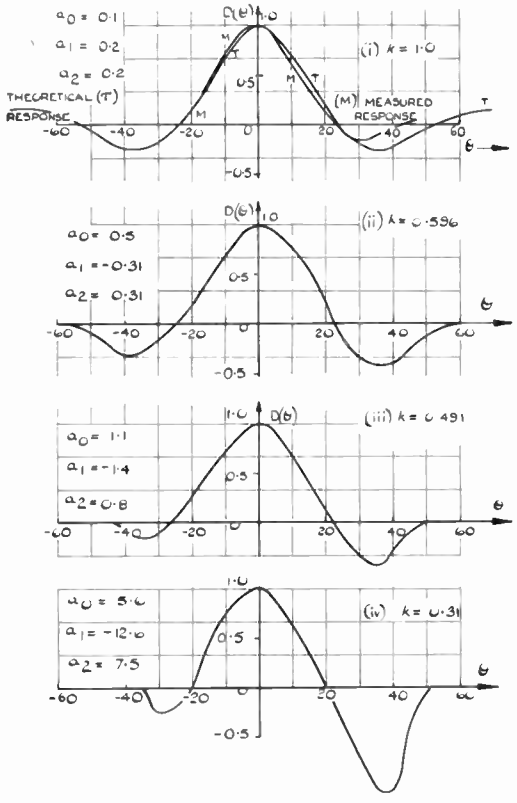


Fig. 20. Superdirective multi-frequency responses.

for all four arrays is shown in Fig. 20(i). The measured responses, which have been normalized, indicate that as k is reduced the beamwidth remains approximately constant as the two receiving elements have their spacing reduced by a ratio of over three-to-one; the side-lobes are unsymmetrical indicating that the phase-shift circuit was not sufficiently accurate; the error is particularly apparent when k is small. Also as k was reduced the peak output was reduced and fluctuations in the output increased, becoming 50% of the peak at $\theta = 0$ when $k = 0{\cdot}31$. Unfortunately as there was a change of surface state during these tests the increased fluctuations cannot necessarily be attributed to a poorer noise factor. It is clear that a useful degree of superdirectivity has been demonstrated over the observable range, $-40° < \theta < +40°$; but in spite of the absence of mutual coupling with bandpass working, high degrees of superdirectivity are still difficult to obtain, mainly due to the difficulty of controlling the taper coefficients to the required degree of accuracy.

5. Other Aspects of Multi-frequency Arrays

The directional patterns of all the multi-frequency arrays apply only for a single incident wavefront. The pattern obtained in the presence of two or more incident wavefronts cannot be obtained by simple superposition. Similarly the directivity index is not a complete guide to the performance of the array when noise is present.

5.1. *Multiple Wavefronts*

It may be shown[6] that in the presence of M wavefronts the correlator output consists of the addition of M displaced patterns, with maxima corresponding to the M directions, plus $M(M-1)$ cross product patterns; the latter are randomly displaced and cause a confusion which becomes worse as M increases. If there is a relative change in the times of incidence of the wavefronts at the array of at least one fundamental period, the cross-product terms may integrate out; this is only possible when the fluctuation period is short compared with the overall integration time.

Fluctuating multiple paths from the acoustic source to the array may be considered to cause a disturbing noise on a steady level at each frequency, so that the taper function is variable. If these changes in level at each frequency are completely correlated then the directional response of the array is unaltered except for a change in level as for a multi-element array. However, if the changes are not correlated between frequencies, then the taper function and pattern are variable. Generally one would expect the cross-correlation coefficient of the amplitudes of two frequencies to be a decreasing function of frequency spacing.

5.2. *Signal/noise Performance of Multi-frequency Arrays*

The bandwidth of the signal modulation on each harmonic will be restricted to $\pm \delta\omega$, which is therefore the bandwidth for which noise is significant. If the input signal/noise ratios in each band and each channel are R it may be shown[3, 6] that the total output signal/noise ratio R_{ot} is given by

$$R_{ot} = \frac{R^2\sqrt{m}}{\sqrt{(\frac{1}{2}+R^2)}}$$

where m is the number of harmonics used.

Thus when $R > 1$ the array gives an improvement of $\sqrt{m}$ over that of a single band; when $R \ll \frac{1}{2}$, $R_{ot} \rightarrow R^2\sqrt{2m}$, there being a rapid decrease in the signal/noise ratio improvement. Direct comparison with other types of array is difficult because physical dimensions are not comparable, and another basis of comparison is necessary. A linear array of $2m$ elements may be chosen to give the same directional pattern against a single source as the multi-frequency array; for the same modulation bandwidth $\pm\delta\omega$ the two arrays will have the same channel-bandwidth capacity $4m\,\delta\omega$. The r.m.s. output signal/noise ratio of a $2m$ element uniform linear array, for which the input signal/noise ratio is R in each channel, is $R\sqrt{2m}$, providing the noise is uncorrelated at each element.

The multi-frequency array is thus seen to be inferior by a factor $\sqrt{2}$ at high values of R and by a factor $1/R$ for low values of R.

6. Conclusions

The theoretical basis of the space-frequency analogy and its limitations have been described. The usefulness of the analogy for the calculation of the directional patterns of wide-band, two-element, correlation arrays has been indicated and confirmed using band-pass operation. It has also been pointed out that the directional response is not a complete guide to the performance of the array when multiple wavefronts or noise are present. As with many processing systems employing non-linear devices, the signal/noise performance is poorer at low input signal/noise ratios. Owing to this, the performance of an echo-ranging system using a two-element array receiving multi-frequency echoes would be worse than a conventional, but larger, array with the same directional pattern for single frequency echoes. The size restriction on a receiving array would need to be severe before the reduced signal/noise ratio would be ignored, especially in a sonar system, where reverberation limits detection at all ranges; in a radar system the main effect and objection would be the reduced range. It might be advantageous to use the correlation array as a short range system with good angular resolution, for its size. However its range resolution is poor, owing to the reduced modulation bandwidth, and its ability to discriminate between multiple targets at the same range but different bearings is also poor; thus the reduced beamwidth is of little consequence, unless a single target at high signal/noise ratio is known to exist, as in a tracking system, or unless the bearing of a transmitting source is sought.

Where a radio communication transmission uses frequency-division multiplex, the multiple carriers may be filtered from their modulations at two receivers and used as the multiple frequencies for the accurate bearing determination of the transmitter. This may be useful for a navigational fix or for tracking a satellite, especially with a split-beam response in conditions of fading.

An underwater acoustic application employing one-way transmission would be for a ship in fog to transmit a multi-frequency signal, warning other ships as well as enabling them to determine its bearing; this is the underwater analogue of a fog-horn and the acoustic analogue of a lighthouse consists of transmitters sited by rocks, wrecks, sandbanks and other navigational hazards; in fog these transmit a suitable wide-band signal, from which a ship could obtain a fix, when within warning range.

Besides the straightforward multi-frequency arrays, good agreement with theory was achieved for split-beam arrays, for arrays receiving random noise and for the lesser known hybrid clipped array, which has possibilities for electronic scanning. Though multi-frequency superdirectivity has been demonstrated and in spite of the absence of mutual coupling effects, high degrees of superdirectivity are still difficult to obtain owing to the accuracy of taper coefficients required.

7. Acknowledgments

The work described above was performed in the Department of Electrical Engineering at the University of Birmingham. The author would like to record his thanks to Professor D. G. Tucker for encouragement and facilities and to Dr. V. G. Welsby for many helpful discussions.

8. References

1. W. E. Kock and J. L. Stone, "Space-frequency equivalence", *Proc. Inst. Radio Engrs*, **46**, pp. 499–500, February 1958 (Letter).
2. W. E. Kock, "Related experiments with sound and electromagnetic waves", *Acustica*, **9**, p. 227, 1959.
3. V. G. Welsby, "Two-element aerial array", *Electronic Technology*, **38**, pp. 160–3, May 1961.
4. V. G. Welsby, "Electronic sector-scanning array", *Electronic Technology*, **39**, pp. 13–18, January 1962.
5. D. G. Tucker, "Space-frequency equivalence in directional arrays", *Proc. Instn Elect. Engrs*, **109C**, pp. 191–7, 1962 (I.E.E. Monograph, No. 479E, November 1961).
6. B. S. McCartney, "Sonar Receiving Systems and Arrays", University of Birmingham, Ph.D. Thesis, April 1963.
7. R. H. Wilcox, "Crystal Diode-Ring Multipliers", Naval Research Laboratory, Washington, D.C., N.R.L. Rept. No. 4385, June 1954.
8. D. G. Watts, "A general theory of amplitude quantization with applications to correlation determination", *Proc. Instn Elect. Engrs*, **109C**, pp. 209–18, 1962 (I.E.E. Monograph, No. 481M, November 1961).
9. V. C. Anderson, "Digital array phasing", *J. Acoust. Soc. Amer.*, **32**, pp. 867–70, 1960.
10. D. G. Tucker, "Signal/noise performance of superdirective arrays", *Acustica*, **8**, pp. 112–6, 1958.
11. S. A. Schelkunoff, "A mathematical theory of linear arrays", *Bell Syst. Tech. J.*, **22**, pp. 80–107, January 1943.

Manuscript received by the Institution on 1st November 1963. (Paper No. 923)

Contributions to the discussion on this paper appear on page 144.

DISCUSSION

Under the chairmanship of Mr. W. K. Grimley

Dr. M. I. Skolnik: Although you have not defined what you mean by a superdirective array, when applied to single-frequency conventional antennas, superdirectivity has a precise formulation and connotes a high-Q aperture with narrow bandwidth and large circulating currents. It is not clear that the multi-frequency array is superdirective in the same sense as the conventional antenna. It would seem more appropriate to call this array by some other name, perhaps pseudo-superdirective, so as to not confuse the two.

Dr. McCartney (*in reply*): Superdirectivity has deliberately not been defined in the paper since it is believed that the change from a normal to a superdirective array as k is reduced is a gradual one. If a precise demarcation was necessary perhaps the point at which reversed phase connection of alternate elements became necessary would be the most suitable. Whilst narrow-band operation is inherent in a superdirective array, for a receiving array this does not always mean high-Q circuits with large circulating currents, as stable buffer amplifiers may follow the array elements prior to the introduction of the superdirective taper. However, a superdirective transmitting array cannot avoid being a high-Q aperture with large circulating currents. The aperture improvement ratio, noise-factor and narrow bandwidths are properties of superdirective arrays assuming ideal, independent, point elements, whilst high-Q circuits and mutual coupling between elements are practical—albeit important and often unavoidable—difficulties associated with the physical realization of the array.

Since reciprocity in the normal sense does not apply to multi-frequency arrays, single-frequency array theory may only be applied to multi-frequency receiving arrays via the analogy.

For the so-called 'superdirective' multi-frequency array the aperture improvement ratio applies, as the spacing between the two elements is reduced for the same overall multi-frequency bandwidth; also the modulation bandwidth is reduced and the noise factor worsened, so that the same basic properties as a single-frequency superdirective array are obtained. However, the practical problems involved in the physical realization take a different form. Providing such arrays are given the full title of 'multi-frequency superdirective arrays' confusion should not arise, whilst strictly all multi-frequency arrays are 'pseudo', whether superdirective or not.

Mr. E. D. B. Shearman: Would Dr. McCartney please say a little more about the interesting idea that the poor noise factor of a superdirective array is attributable to the large side-lobes which occur outside the range of real angles.

Dr. McCartney (*in reply*): The relative phase angle at two elements between noise components arising from the medium cannot exceed $k\pi$, whilst for noise components which arise in the receiving elements prior to the taper formation there is no restriction, i.e. $-\pi \leqslant \psi \leqslant \pi$. Noise factor applies to the latter type of noise. The contribution in the output due to each noise component will be proportional to $D(\psi)$. Thus the total r.m.s. output noise voltage will be proportional to

$$\left(\int_{-\pi}^{+\pi} D^2(\psi) \, . \, \mathrm{d}\psi \right)^{\frac{1}{2}},$$

whilst the peak output signal voltage is $D(0)$. Thus as the side-lobes outside the range of ψ corresponding to real angles increase the noise factor increases.