Founded 1925

Incorporated
by Royal Charter 1961

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

VOLUME 42 No. 5 MAY 1972

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

Radio Receivers and Associated Systems

University College of Swansea, 4th to 6th July, 1972

IT was from the need to construct radio receivers that electronics had its origin. Today the radio receiver is numerically the commonest item of electronic equipment throughout the world. In the domestic field alone, the first cost of radio and television receivers in use in Europe must approach £1000M. Add to this investment in professional, aircraft and marine use, radio navigation, mobile and personal radio, and it is clear the scale on which radio receivers are used. It was, however, less these considerations than awareness that receiver technology is changing more rapidly now than at any time which prompted the Institution to organize (jointly with the I.E.E. and I.E.E.E.) the forthcoming conference under the above title.

For years the design of radio receivers had been on a plateau. The perfection of the superheterodyne led to a long period in which there seemed to be few outstanding problems. Solid-state active devices and integrated circuits have completely changed this situation. Because receivers can now be designed orders of magnitude more complicated than before, with improved reliability and economic cost, a new approach becomes possible.

The first major change was the use of a synthesizer instead of a free tuning local oscillator. Whole new areas of scientific enquiry were opened up, concerned with synthesizer noise and its consequences, and rapid development is now occurring. Similarly, it is now possible to consider adaptive receiver and aerial systems, which change their characteristics to exploit the particular properties of the signal received. Microelectronics also makes practicable new direct conversion receiver configurations, based on low cost complexity and excellent matching between circuits on a chip, which bid to break the domination of the superheterodyne.

Another current pressure on receiver technology arises from spectrum congestion. An aspect of the over-utilization of the finite resources of our planet is the increasing growth of congestion on radio bands, making it difficult even to maintain existing standards of service. Receivers for more efficient spectrum utilization will be of increasing importance, as will advances in elimination of spurious receiver responses, and all other techniques aimed at improving reception in congested conditions.

These and many other topics will form the matter of the forthcoming conference. Without radio the civilization that we know could not exist, and the electronics industry would not have been born. An industry in a difficult economic climate should not forget that radio, in all its forms, is still its largest market.

WILLIAM GOSLING

Contributors to this issue

Mr. R. E. F. Bywater graduated from Battersea College of Technology, London, with a B.Sc. degree in electrical engineering in 1964. After spending a further year reading computer engineering, he joined the Computer Systems Development Group of ICL, then English Electric Computers. Currently, he is a Lecturer in electrical engineering at the University of Surrey with research interests in special purpose computers.

Professor W. F. Lovering graduated from Birmingham University in 1935 with an honours degree in electrical engineering and was awarded the degree of M.Sc. for extra-mural research in 1945. After experience in the radio industry he held lecturing posts at Nottingham University, S.W. Essex Technical College, University of New South Wales, and Imperial College, before joining Battersea College of Technology (now Surrey University); he was appointed to chair in 1964. Author of 'Radio Communications' and of several papers on circuit design, his current interest is in special-purpose computers. Professor Lovering is a member of Joint I.E.E.E.–I.E.R.E.–I.E.E. 'Travelling Scholarships' Committee.

Professor D. P. Howson (Fellow 1969, Member 1961) who is a graduate of Bristol University, gained his industrial training when working as a design engineer with Standard Telephones & Cables from 1952 to 1957. In 1958 he received an M.Sc. degree in information engineering from Birmingham University and was then appointed a lecturer in the Department of Electrical and Electronic Engineering. In 1967 he was appointed to a chair of Electrical Engineering at the University of Bradford and is currently Chairman of the Postgraduate Board of Studies in Electrical and Electronic Engineering. Professor Howson has been the author of several technical papers in the *Journal* and of the book 'Mathematics for Electrical Circuit Design'. He has served as Chairman of the West Midland Section and on the Examinations Committee and is now a member of the Education and Training Committee.

Dr. Joseph Helszajn, who was born in Brussels, received the Full Technological Certificate of the City and Guilds of London Institute while at the Northern Polytechnic in 1955, the M.S.E.E. degree from the University of Santa Clara, California, in 1964, and the Ph.D. degree from the University of Leeds in 1969. He has held a number of positions in the microwave industry and from 1964-1966 was a product line manager at Microwave Associates Inc., Burlington, Mass. In 1968 he received the Insignia Award of the City and Guilds of London Institute. Since 1968 he has been working as an independent consultant and is also currently a Senior Research Fellow at Heriot-Watt University, Edinburgh. He is the author of some 25 papers, 10 of which have been published in the Institution's *Journal*, and a book on microwave ferrite engineering.

Mr. H. W. Bryan was educated at Handsworth Grammar School after which he worked at the General Electric Company plant at Witton. He then joined Marconi's Wireless Telegraph Co. working on radar display equipment during which time he obtained his Higher National Certificate. In 1962 he joined the Space Research Department at Birmingham University and has since worked on circuit development for a number of space projects including the *Ariel III* and *Prospero* satellites.

Mr. J. Wall served in the Royal Corps of Signals from 1946-59, and then joined the G. W. Division of the Armstrong Whitworth Aircraft Co. at Coventry, where he worked on the development of missile power supplies. After obtaining his Higher National Certificate he led a team on telemetry and instrumentation. He joined the Space Research Department at Birmingham University in 1968 where he has worked on the design and development of the electron density experiment for the *Ariel IV* satellite.

Mr. J. H. Wager joined the Electron Physics Department of Birmingham University in 1949 following a period of Army service. When space research began at the University in 1955 he became responsible for the design and construction of the instrumentation used. He now leads the Instrumentation Group in the Department of Space Research which, in addition to working on an extensive rocket research programme, has developed equipment for the *Ariel I* and *IV*, *FRI* and *Prospero* satellites.

U.D.C. 681.332.64

A Programmable Extended Resolution Digital Differential Analyser

R. E. H. BYWATER, B.Sc.(Eng.)*

and

Professor W. F. LOVERING,
M.Sc., M.I.M.C., C.Eng., F.I.E.E.*

SUMMARY

A digital differential analyser is described which features the economical use of parallel arithmetic, programmable interconnexions and internally-scaled integrators. It is suitable for connexion to a general-purpose digital computer to form a hybrid facility and as such would be capable of solving dynamical problems and simulating systems. The use of carry-save techniques and a 'trapezoidal' integration algorithm make real-time simulation possible at an acceptable precision.

* *Department of Electronic and Electrical Engineering, University of Surrey, Guildford, Surrey.*

List of Symbols

Z	integral	register
Y	integrand	register
R	integral residue	register
P	potentiometer fraction	register
$H_{(i)}$	value of H at time i	
$H_{(0)}$	initial condition (I_0) of H	
ΔY, ΔZ	increment of Y, Z etc.	
$\Sigma\Delta Y$	summed increments for Y	
t	time (iterations)	
ΔX	increment of independent variable	
$S(H)$	sign of H	
ε	round-off error due to variable discretization	
ω	sine wave generator angular frequency	
ϕ	arbitrary phase angle	
ξ	integrator slew rate	

1. Introduction

There are many problems, for example, in the real-time simulation of continuous physical processes, where the solution rate of a general-purpose digital computer (g.p.d.c.) is inadequate. This stems, in part, from the fact that the g.p.d.c. uses sequential processing methods to obtain a solution. Although the solution time may be reduced by taking wide steps in conjunction with such methods as the Runge-Kutta integration procedure, the improvement possible is limited by the solution error/instability which may be tolerated.

Although the use of special digital integrators was suggested in the 1950s, the cost and speed of available circuits offered little advantage.[1-4] The situation has changed with the availability of modern microcircuit elements, and it is now possible to build differential analysers (d.d.a.) to considerable advantage. These devices can make use of special-purpose logic to effect substantial performance gains over g.p.d.c.s.[5, 6] The use of digital integrators obviates some of the basic difficulties of analogue methods. It is possible to include logic decisions in a program. Such integrators may be simply interfaced with a digital computer to make a powerful hybrid facility, and are programmable without needing special analogue switches or servo-potentiometers.

This paper outlines the design of a d.d.a. which is able to use currently available components. In particular, the integrator is described together with suitable, programmable, interconnexion techniques.

2. D.D.A. Principles

2.1. *The Digital Integrator*

The basic form of a digital integrator is illustrated in Fig. 1. The inputs are increments in the values of the variables X and Y. A 'clock' signal is used to control the operation. At each clock pulse the inputs are interrogated. The increment of ΔY $(+1, 0, -1)$ is added to the previous value of Y and stored in a register. If the ΔX increment is positive, the contents of the Y register are added into the contents of the R register. If the ΔX increment is negative, Y is subtracted from the

contents of the R register; if the ΔX increment is zero, the R register contents are unaltered. The R register thus accumulates the integral increments $Y.\Delta X$. As the increment information can take only three states, it is known as the ternary transfer system. When the total in the R register reaches or exceeds some predetermined value $\pm M$ $(=2^n)$, an output signal is produced and the number M is subtracted from the register contents. Each output signal then represents an increment $\pm M$ $(=\Delta Z)$ of $Y.\Delta X$, that is, $\Delta Z = Y.\Delta X$.

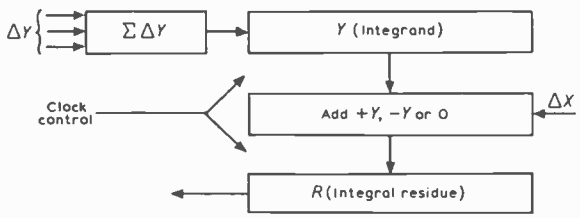


Fig. 1. Basic digital integrator.

In the simplest form, the R register is incremented only according to the current values of Y and ΔX. By the use of additional stores, the integration algorithm may be made to take account of the rate of change of Y and to introduce a corrective element. In this case, the manner in which the R register is incremented is controlled by the current and previous values of ΔY and ΔX. For simple time-dependent equations, the ΔX input is provided by a clock. The ability to integrate with respect to any variable, however, adds to the power of the d.d.a.

In an analogue computer, multipliers and sign reversers are needed. These are not necessary in a d.d.a. since

$$xy = \int x\,dy + \int y\,dx$$

can be simply programmed, obviating the need for a separate multiplier. The complement of a digital number is always available, and this also makes special sign reversers unnecessary.

At the expense of more electronic gates, the ΔZ signals may be obtained as binary words representing a range of values for the ΔZ increment. If this is done then, to solve a differential equation, subsequent integrators must be capable of accepting ΔY increments in the form of binary words. If a word ΔZ is used as a ΔX input, then a binary multiplication is needed to form $Y.\Delta X$.

The effect of using 'word ΔZs' is to more closely define the value of an output increment. Thus, the difference between the true value of the increment and its quantized representation, is reduced. For every extra bit that is used to form ΔZ, the average per-step round-off error is halved.

A d.d.a. being an assemblage of digital integrators and other special-purpose logic elements (function generators, and the like), must have these elements interconnected in various ways to solve different problems. This is the basis of hardware programming. If simple transfer systems are used to interconnect these elements, i.e. ternary, then the interconnexion logic will be simple and inexpensive. However, use of word transfers increases the cost about in proportion to the number of bits transmitted so that an engineering compromise is needed between solution accuracy and machine cost.

2.2. *Program Patching*

Connexion between analogue integrators needs only a single wire which then can convey an 'infinity' of voltage levels. Digital transmission, however, requires either one wire per bit of information or the data may be serialized and transmitted a bit at a time along a single wire.

If the output of a digital integrator is serialized, considerable time may be used in communicating integral outputs. For, say, 16- or 24-bit machines, this could be a large proportion of the total machine iteration time. However, the system does have the merit of simplicity and low cost.

The alternative, that is, parallel data transmission, is much faster, but somewhat more costly. However, the higher iteration rates that result allow simple integration algorithms to be used without loss of accuracy. The savings on algorithm complexity can do much to offset the greater cost of interconnexions. Furthermore the simpler algorithms tend to have a much more predictable performance.

It seems, therefore, that the best way to achieve a very high solution rate (as required for real-time simulation, etc.) is by parallel arithmetic techniques together with an efficient, interconnexion method.[7]

This paper describes a method of obtaining reasonably fast interconnexions which is not costly. The method outlined is a parallel-bit, serial-word version of a total interconnexion topology. By transmitting a certain portion of a unit's output (integral), it represents an engineering compromise between a high-cost total integral system and the low-cost ternary transfer method.

3. Overall System

The d.d.a., as built, is a parallel-arithmetic, 12-bit analyser capable of interconnecting 16 units (integrators, etc.) and has an iteration period of about 3 μs (min). Of this period, 1 μs is spent in the machine cycle, proper, the other 2 μs in effecting the inter-unit connexions.

The integrators implement the trapezoidal correction method on an Euler prediction, and have built-in potentiometers for scaling the output of each integrator over a wide range. Integration can be with respect to time or the output of any one unit. All unit outputs are transmitted as 4-bit rounded increments.

4. Integrator Design

A digital integrator may be considered as a device for integrating (or summing) quantized, sampled input data, so producing a quantized output. This is equivalent to finding the area under the graph of Fig. 2.

The actual input as seen by the integrator is the set of points only, and it is apparent that some uncertainty exists (and errors may be induced) by the amplitude and time discretization. However, the integrator has only these discrete values to work with, and to what good effect it uses these points, either singly or severally, for each integration step, will determine the precision of its output.[8]

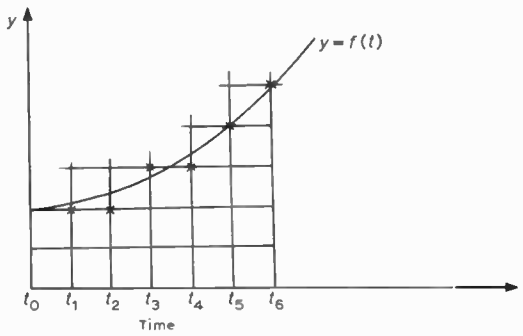


Fig. 2. Time and amplitude discretization of a continuous input function.

The integrator implements an Euler prediction followed by a retrospective (first-order) trapezoidal correction (Fig. 3). Thus on step (i), the current ordinate y is $Y_{(i)}$ $(= Y_{(i-1)}+\Delta Y_{(i)})$. The integral is thus $P.Y_{(i)}.\Delta t_{(i)}$ where P is a fractional scaling factor (potentiometer fraction). Using incremental notation for input and output:

$$\left.\begin{aligned} Y_{(i)} &= Y_{(i-1)}+\Delta Y_{(i)} \\ \text{or} \\ &= Y_{(i-1)}+\Sigma\Delta Y_{(i)} \text{ (if the integrator is to sum several inputs)} \\ \text{and} \\ \Delta Z_{(i)}+R_{(i)} &= R_{(i-1)}+P.Y_{(i)}.\Delta t_{(i)} \end{aligned}\right\} \quad \ldots\ldots(1)$$

where $\Delta Z_{(i)}$ is an increment of integral

$R_{(i)}$ is the untransmitted residue of integral of lesser significance than $\Delta Z_{(i)}$.

The post correction (PC) using the new ordinate increment $\Sigma\Delta Y_{(i+1)}$ is a triangle of area:

$$\text{PC} = \tfrac{1}{2}\Sigma\Delta Y_{(i+1)}.\Delta t_{(i+1)}. \quad \ldots\ldots(2)$$

Equations (1) and (2) may be recast as follows:

$$\left.\begin{aligned} Y_{(i)} &= Y_{(i-1)}+\Sigma\Delta Y_{(i)} \\ \Delta Z_{(i)}+R_{(i)} &= R_{(i-1)}+P.\Delta t_{(i)}\{Y_{(i)}+\tfrac{1}{2}\Sigma\Delta Y_{(i)}\} \end{aligned}\right\} \quad \ldots\ldots(3)$$

It has been pointed out[9] that equation (3) may be recast so as to avoid additions of quantities involving long registers i.e. Y, R, as follows:

$$\left.\begin{aligned} Y_{(i)} &= Y_{(i-1)}+\tfrac{3}{2}\Sigma\Delta Y_{(i)}-\tfrac{1}{2}\Sigma\Delta Y_{(i-1)} \\ \Delta Z_{(i)}+R_{(i)} &= R_{(i-1)}+P.Y_{(i)}.\Delta t_{(i)}. \end{aligned}\right\} \quad \ldots\ldots(4)$$

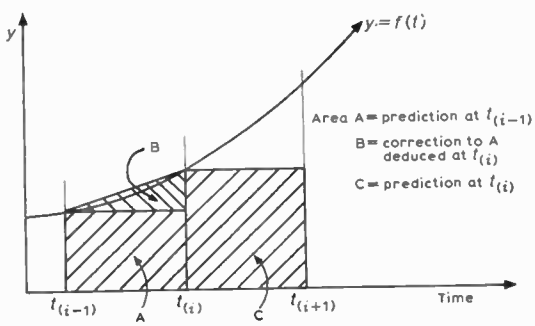


Fig. 3. Euler integration of $y=f(t)$ with retrospective trapezoidal correction.

Although one more addition is involved, the two additions of $\Sigma\Delta Y$ are of very short words. Furthermore, as will be shown in the hardware description, carry-save techniques can be more effectively implemented.

5. Integrator Hardware

The system is completely built around currently available 7400 series TTL medium speed integrated circuits. Lately, several new circuits have come on to the market which could improve packaging and performance.

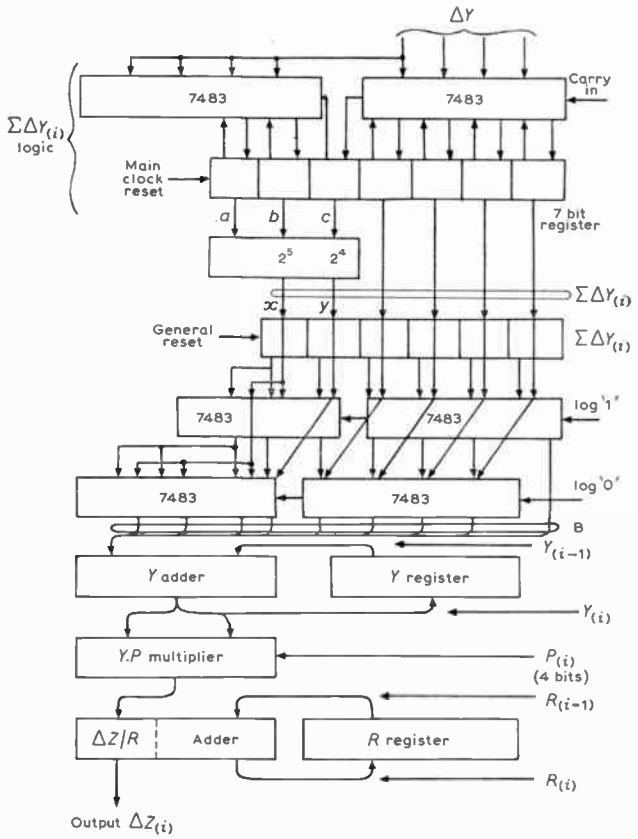


Fig. 4. Integrator dependent variable input logic.

In describing the integrators, it is difficult to be precise about the point where interconnexion logic gives way to integrator logic, particularly as the integrators, as built, can accept up to four inputs. However, the integrator will be described from the point of receipt of $\Sigma\Delta Y_{(i)}$ prior to staticization. (Partial sums of $\Sigma\Delta Y_{(i)}$ will have been stored, as formed, in a register, and the final value will be the sum of the penultimate final sum and the last entry.)

5.1. *Dependent Variable Logic*

Figure 4 shows the logic for the $\Sigma\Delta Y_{(i)}$ accumulator and the associated summing logic. The design is arranged to make the best use of available microcircuits both from the cost and performance viewpoints. Thus despite some logic redundancy, the quadruple full adder (7843) frequently occurs. The system built uses four bits (2's complement notation) to represent machine increments $(\Delta Z, \Delta Y, \Delta X)$ giving a range $-8 \leqslant \Delta Z, \Delta Y, \Delta X \leqslant +7$. An integrator fan-in of four thus causes $\Sigma\Delta Y$ to lie in the

range $-32 \leqslant \Sigma\Delta Y \leqslant +28$, i.e. 100000 through 011100, i.e. six bits.

Carry-save techniques are used in the accumulator. The carry-out from the least significant (1.s.) quadruple adder is stored in a single flip-flop, to be applied to the carry-in on the next clock pulse. After the final interconnexion clock pulse, the last 'saved' carry is assimilated into the intermediate sum of digits of the ms part of $\Sigma\Delta Y_{(i)}$. The assimilator (Fig. 4) uses a logic array implementing:

$$\left.\begin{aligned} x &= a(\overline{bc}) + \bar{a}bc \\ &= a \oplus (bc) \\ \text{and} \quad y &= b \oplus c = b\bar{c} + \bar{b}c \end{aligned}\right\} \quad \ldots\ldots(5)$$

Carry-save allows selection of ΔYs to take place at about 8 MHz, despite the adder logic delays. The final carry assimilation delay is about 20 ns. This method compares with a typical, estimated, selection rate of 4 to 5 MHz if carry-save is not used. The six bit $\Sigma\Delta Y_{(i)}$ is added, suitably weighted, to $\Sigma\Delta Y_{(i-1)}$ to form $B = \frac{3}{2}\Sigma\Delta Y_{(i)} - \frac{1}{2}\Sigma\Delta Y_{(i-1)}$. The whole expression is formed in two stages by two cascaded adders. Two possible reduced versions of B are

$$\left.\begin{aligned} &B = 2\Sigma\Delta Y_{(i)} - \tfrac{1}{2}\Sigma\Delta Y_{(i)} - \tfrac{1}{2}\Sigma\Delta Y_{(i-1)} = d+e+f \\ &\text{or} \\ &B = \Sigma\Delta Y_{(i)} + \tfrac{1}{2}\Sigma\Delta Y_{(i)} - \tfrac{1}{2}\Sigma\Delta Y_{(i-1)} = g+h+f \end{aligned}\right\} \quad \ldots\ldots(6)$$

for which a total of six combinations of processing exist, i.e. $(d+e)+f$, $(e+f)+d$, etc. The method chosen is $(h+f)+g$ as it calls for the least logic without impairing the machine performance. The machine, with an integrator fan-in of four and four-bit increments, needs only seven-bit adders at both the first and second stages. Pairs of quadruple full adders are used so that fan-in expansion can take place, if desired.

There is only combinatorial logic between the $\Sigma\Delta Y_{(i)}$ accumulator all the way through to the ΔZ logic. (Fig. 4.) All the register updates, i.e. $Y_{(i)}$, $R_{(i)}$, etc., are formed as 'spin-off' results down the logic chain. This is only slightly more expensive than a system which forms and staticizes intermediate results at each stage, since although adder time sharing cannot take place, the expense of adder input selection logic is eliminated. A system, such as this, which uses a number of logically cascaded adders, gives a better performance than a time-sharing adder system. The total carry propagation through all the adders is only slightly greater than that through one. In addition, delays in adder selection logic are obviated.

When the iteration has finished, the $\Sigma\Delta Y_{(i-1)}$ register is updated with the current value of $\Sigma\Delta Y_{(i)}$ ready for the next iteration. The same clock signal is used for updating $Y_{(i)}$, $R_{(i)}$, and $\Delta Z_{(i)}$. At each iteration $(\frac{3}{2}\Sigma\Delta Y_{(i)} - \frac{1}{2}\Sigma\Delta Y_{(i-1)})$ is added to the current contents $(Y_{(i-1)})$ of the Y (ordinate) register. This is done with a separate set of three 7483's (Fig. 4) so as to be consistent with the combinatorial method mentioned.

The $\Sigma\Delta Y_{(i-1)}$ register has a reset line which is actuated when the rest of the machine is reset. Ideally, at time $t = 0$, the $\Sigma\Delta Y_{(i-1)}$ register would be preset to the slope of Y. In practice, this initial slope is not known, unless a 'dummy' iteration has been carried out with guessed values of $\Sigma\Delta Y_{(0)}$ to give a reasonable estimate. (This would be somewhat on the lines of the Adams-Moulton predictor-corrector scheme.)

However, for iterative computing and hill-climbing, it is desirable, at least, to have repeatable initial contions for all parts of the machine. Resetting the $\Sigma\Delta Y_{(i-1)}$ register to zero at $t = 0$ satisfies the condition of repeatability. The error induced is not very great as it only directly affects the first iteration.

The Y register is capable of being overloaded (overflow), and suitable logic is appended to detect this. As only addition of $Y_{(i-1)}$ and B can take place, the overflow (F) is simply described by

$$F = 1 \quad \text{if} \quad [S(Y_{(i)}) \not\equiv S(Y_{(i-1)})] \wedge [S(Y_{(i-1)}) \equiv S(B)] \quad \ldots\ldots(7)$$

The Y register can be foreshortened from 12 to 8 bits if higher speed, lower accuracy computation is required. Separate overflow logic is sited at the appropriate points on the $Y_{(i)}$ adder, to cater for both situations.

5.2. *Scaling Logic*

In order to reduce the number of possible machine elements, and hence interconnexion logic, the output of each integrator should be scalable by some constant. This is equivalent to hardwiring a potentiometer to each integrator output in an analogue system. Furthermore, it was decided that, as machine expansion might demand that more than one scaled value should be made available, the logic design should be capable of such an expansion merely in the form of added logic—not a complete redesign. Because scaling involves the multiplication of one number (the potentiometer fraction, held in a register) by the current contents of the Y register, it would appear that a full multiplier is necessary. However, despite the fact that the machine contains 12-bit Y and 8-bit P fields, only a 12-bit product is required, of which 4 is used for ΔZ. Thus P can be reduced to 4 bits provided these 4 bits are chosen (a) to average out to the desired 8-bit fraction over a large number of iterations and (b) the average P at any time is as nearly rounded to the desired fraction as possible.

The P register is constrained to perform in the manner outlined by adding a second register and adder which accumulate the amount by which the currently used 4-bit version of P differs from the desired fraction. The operation is not unlike that of the $\Delta Z/R$ logic. Thus P is added, once per iteration, into its auxiliary P' register and the most significant 4 bits of the sum is used as the multiplier for forming $Y_{(i)}.P$. The less significant residue is staticized in P' to be accumulated with P on subsequent iterations. (Fig. 5.)

To form a fraction 23/32 from a 4-bit multiplier (m), i.e. $m = k/8$ where $-8 \leqslant k < 8$, requires m to assume values of:

$\frac{3}{4}, \frac{3}{4}, \frac{5}{8}, \frac{3}{4}, \frac{3}{4}, \frac{3}{4}, \frac{5}{8}$ ... etc. at $t = i, i+1$, etc.

The limit at which there is any meaning in the precision of P is determined by the rate at which $Y_{(i)}$ can change from its initial value at (i) while P is being run through a

cycle of values whose mean corresponds to the desired fraction (in this case 23/32 in 4 cycles). This will be determined by the machine accuracy required, the integrator fan-in, and the significance at which B is added to $Y_{(i-1)}$.

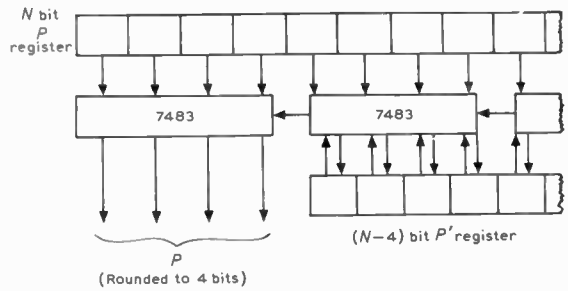


Fig. 5. Scaling factor generator.

Before computation starts, P' is preset to $+\frac{1}{16}$ so that the top 4 bits of P' are suitably rounded to the nearest $\frac{1}{8}$ quantum to give the best multiplier for any given step. For the example cited, the register contents are:

$P = \frac{23}{32}$

P'	$\frac{1}{16}$	$\frac{25}{32}$	$\frac{24}{32}$	$\frac{23}{32}$	$\frac{26}{32}$	etc.	—
$P'(ms4)$	0	$\frac{6}{8}$	$\frac{6}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	etc.	—
$P'(ls)$	$\frac{1}{16}$	$\frac{1}{32}$	0	$\frac{3}{32}$	$\frac{1}{16}$	etc.	—
t	0	1	2	3	4	5	6

The multiplier is fed with $Y_{(i)}$ on one side (12 bits) and P' on the other (4 bits). It was found that an economical multiplier could be formed from three 7483's by encoding P' such that 2 bits at a time multiplication was carried out on either side of the adder, i.e. either ± 1 or $\pm\frac{1}{2}$ times $Y_{(i)}$ was set to one side and either $\pm\frac{1}{8}$ or $-\frac{1}{4}$ $Y_{(i)}$ to the other. Thus the selection table for all P' becomes:

$P' \times \frac{1}{8}$	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8
L.H.S.	−1	−1	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1
R.H.S. ($\times\frac{1}{8}$)	0	1	−2	−1	0	1	−2	−1	0	1	−2	−1	0	1	−2	−1	0
Adder carry-in	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓

For $P' = -\frac{3}{4}$ and $-\frac{5}{8}$, a carry-in of significance 2^1 (not 2^0) is required as both left- and right-hand factors to the adder are negative. In fact, occasional bottom bit errors thus introduced are of little importance compared with those due to per step round off error (relative significance to round-off is 2^{-q} where q is the length of the R register).

5.3. *Output Logic*

The scaled value of $Y_{(i)}$ is applied to the R register in a conventional way to form the sum $(R_{(i-1)}+P.Y_{(i)})$. The 4 most significant bits of this sum form $\Delta Z_{(i)}$, the least significant are stored as a less significant residue in the R register. The updating of this register is under the command of the main clock. As with other incremental systems, the output (ΔZ) can be rounded correctly to the nearest quantum by setting R_0 to $+\frac{1}{2}$ unit quantum before computation starts, thus ensuring that the error due to discretization (ε) of ΔZ lies in the range $-\frac{1}{16} \leqslant \varepsilon < +\frac{1}{16}$ and not $-\frac{1}{8} \leqslant \varepsilon < 0$ as would otherwise occur. The improvement is far better than two-fold as the error is now symmetrical about zero.

If required, the P register, multiplier, R register and ΔZ logic may be repeated several times so as to produce several scaled versions of the integral. The most economical approach is probably to have a scaled and an unscaled version of the output, thereby making only an extra R (ΔZ) adder/register necessary. The integrator may be connected to accept one 4-bit word (e.g. an output ΔZ from an integrator) as a ΔX input. The multiplier is then used to form the product $Y.\Delta X$ (instead of $Y.P$ used in scaling). Thus each integrator may be used to give either scaled time integration *or* unscaled integration with respect to an arbitrary variable.

If it should be required to form $(Y.\Delta X_1 + Y.\Delta X_2)$ this can be effected by two integrators receiving the same ΔY signal and different ΔX signals. The outputs are then summed in the following machine element.

6. Interconnexions

The d.d.a. built is capable of interconnecting any machine element to any other by logic gating. The result is a compact system which not only allows programmed interconnexions, using paper tape or computer control, but also provides the hardware for entering initial conditions to the machine elements.

For economy, a simple interconnexion system is used making use of the time domain to effect all the possible interconnexion paths (Fig. 6). The 4-bit outputs from the 16 machine elements are concurrently staticized in a 4 by 16-bit shift register under main clock command. They are then serially entered, via the shift register, to a 4-wire busbar connecting all the machine elements. Whilst this shifting is taking place, a shift register, sited in each integrator, is circulated. The pattern in this shift register corresponds to those ΔZ which the machine element containing this shift register requires to accept.

Thus a '1' can be used for acceptance of a given increment, '0' for rejection. Necessarily, as many shift pulses are required to circulate the data as there are machine elements.

Two embellishments can be applied to this method.

(i) The first, which has been implemented, is to add a second circulatory shift register to each integrator. This is used to determine whether any given accepted increment should be accepted *per se*, or complemented, i.e. a 0 for original form, a 1 for complemented form. This makes programming much more flexible and economical as inverters are rendered unnecessary. Thus, the same output from a given integrator may be presented in its true form to some integrators and in its complemented (negated) form to others. This technique also makes it possible to restrict the range of P to positive values only without loss of programming flexibility.

(ii) The second possible improvement is to multiple-rank the ΔZ and selector shift registers, say, either in twos or fours and make each only a half or quarter of its original length. No extra logic is required to store the ΔZ or selection/negation patterns, only the ability to add more than one 4-bit increment at a time into the $\Sigma\Delta Y_{(i)}$ accumulator. The extra logic is not very great. However, there is approximately a two or four fold decrease in the interconnexion time. This may be worthwhile in a large d.d.a.

An alternative way of reducing the interconnexion time is by a change of the interconnexion topology. This is discussed in the 'Machine Expansion' section below.

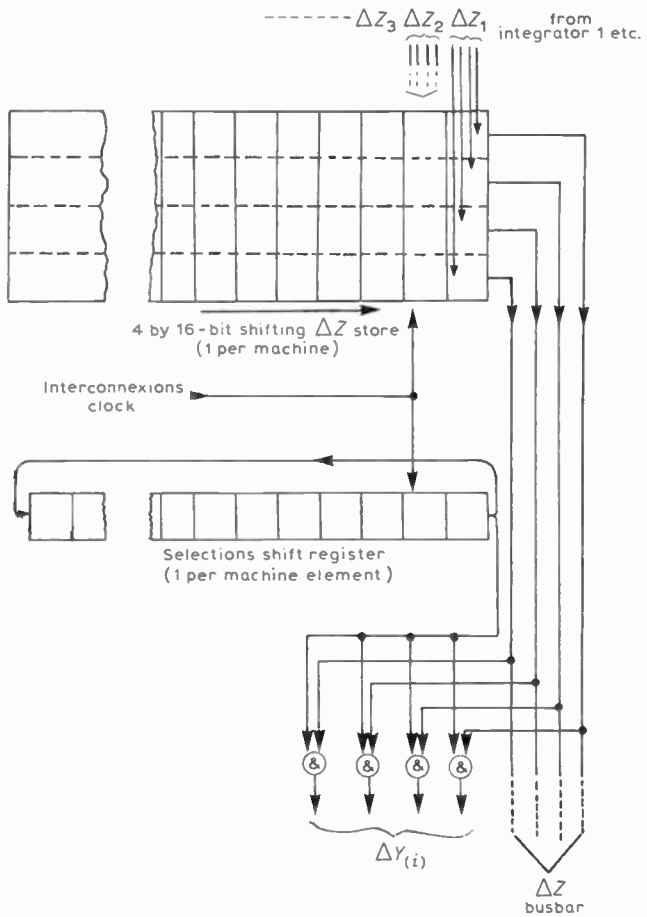


Fig. 6. Interconnexion logic.

7. Input/output

7.1. *Machine Loading*

Since the interconnexion system communicates from a central point to every machine element it may be used for initial loading. The four ΔZ registers together with a fifth 16-bit register are connected during the loading phase as one single 80-bit shift register.

The following initial conditions are serially entered to the 80-bit shift register (Fig. 7):

(i) The scaling fraction P
(ii) The initial value of $Y(Y_{(0)})$

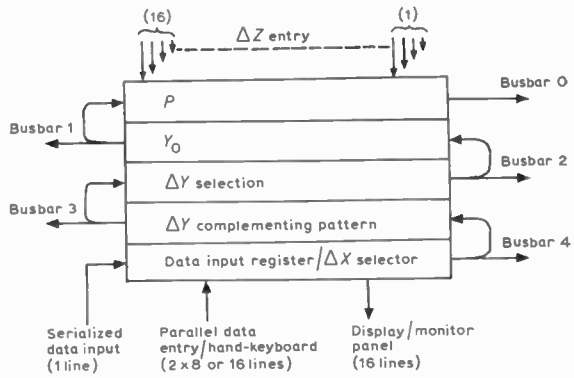


Fig. 7. Data entry/ΔZ store register matrix.

(iii) Interconnexion selector contents
(iv) Interconnexions sign selector contents
(v) ΔX selector contents.

In effect, five 16-bit registers are connected as one long shift register to allow data assembly prior to distribution. The fifth shift register is arranged to accept data either (a) purely serially from a telecommunications line, (b) in parallel from a 1/0 busbar, or (c) separately from panel push-buttons. It is evident that making the ΔZ registers double as data assembly registers has two consequences:

(i) The combined lengths of the shift registers used must be sufficient to hold all the initial conditions for a single integrator. For this machine, the P and Y registers are 8 and 12 bits respectively and the three selectors each of 16 bits. The shift registers must also be individually long enough to hold the 4-bit outputs from each machine element. The machine built, therefore, has a 5 by 16-bit shift register matrix.

(ii) Each machine element must be capable of routing accepted data from the busbars either to the initial conditions registers during loading or the increment input registers during computation. This was achieved by broadcasting the initial conditions serially along the ΔZ busbars to all machine elements and, at the same time, clocking only the element that was to input these data. Although there are only 4 ΔZ bits broadcast by the interconnexion logic during computation, there are 5 busbars so that all the initial conditions can be transmitted in the same way. Machine loading is carried out serially because the frequency of loading compared with that of iterating is so small that the extra logic needed for parallel loading is difficult to justify. Furthermore, the very high gate/pin ratio so far achieved for the machine elements would be largely lost if parallel loading was implemented. The possibility, as i.c. technology advances, of putting complete elements on a chip, is quite evident. However, chip technology is allowing gate counts per chip to rise faster than lead-outs. To some extent this is to be expected, but it does mean that computer sub-systems of ever higher gate/pin ratios are going to be demanded if full use of l.s.i. is to be realized. (There is also a maintenance/reliability penalty

associated with i.c.s of high lead-out counts.) The same arguments tends to make serial interconnexion methods more desirable than parallel ones even if there is a speed penalty.

7.2. *Data Output*

The transmission of data between machine elements is incremental (4 bits). Furthermore the use of the single-step hardware method for implementing the trapezoidal algorithm means that a Y register rarely contains the actual variable y that might require monitoring. (Y corresponds to y only if $\Sigma\Delta Y_{(i)}=0$ whereupon any outstanding $\frac{1}{2}\Sigma\Delta Y_{(i-1)}$ has already been processed.) Thus, if inspection of any given y is required, it is evident that the most desirable way is to assemble the ΔZ in counters which can be addressed like any other machine element. That is, the ΔZ may be accumulated to form Z ($=\Sigma\Delta Z$), see Fig. 8. Data assembly of this sort might be used for outputting information to the general-purpose digital computer or parallel logic/decision systems local to the d.d.a. These might include comparators and limiters and g.p.d.c. interruption generators. Also the output may, via digital-to-analogue (d.a.c.) convertors, monitor solutions on a c.r.o./strip recorder or X-Y plotter.

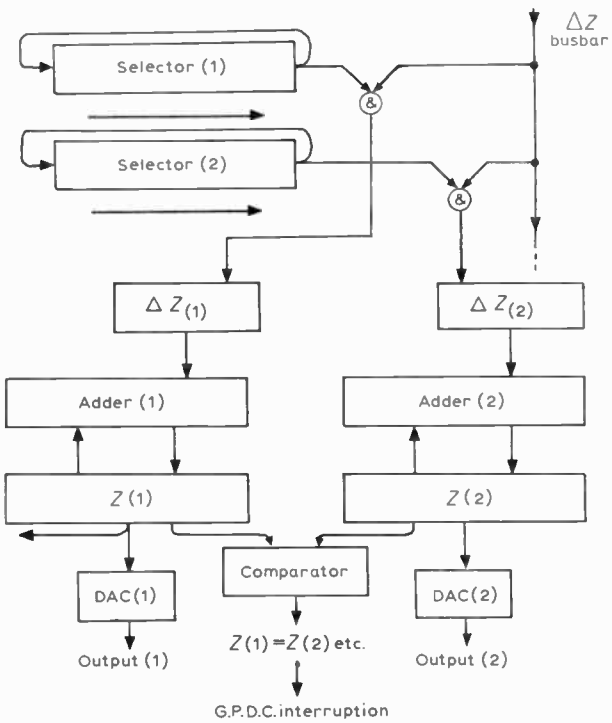


Fig. 8. Addressable d.d.a. monitor logic.

The machine, as it stands, merely makes use of a pair of d.a.c.s to monitor the outputs of two selected integrators. These are wired across at present, for simplicity. However, to provide the extra facilities cited is very straightforward. In particular, providing the outputs to the input/output busbar of a g.p.d.c. allows the latter's peripheral devices to be used, if required. Such a provision, together with a data input facility, can provide completely 'closed shop' hybrid computing devoid of the necessity for analogue computer style patchboards or expensive servo-setting electronics.

8. Performance

8.1. *General Considerations*

The total iteration time of a parallel d.d.a. is a combination of integration and interconnexion time. The former might be termed 'useful' algorithm time, the latter an undesirable overhead. It is desirable to make the interconnexion/integration time ratio as small as possible. This ratio, to a first approximation, is a constant for a given system design, regardless of technology used.

With 7400 series TTL a clock rate of about 8 MHz can be achieved for the interconnexion phase. The limitation lies in the ability of the $\Sigma\Delta Y$ logic to assimilate selected ΔY from the ΔZ busbar. The use of carry-save in the input logic is thus well justified as it represents a strengthening of the weakest link. (The shift registers are capable of conventional operation at about 20 MHz.) The interconnexion period for this 16 element, single rank selection machine is thus

$$16 \times 0{\cdot}125\ \mu s = 2\ \mu s$$

The integration time, which starts from the final selection, is dependent purely on combinational logic, followed by a single clock pulse to staticize the result in the ΔZ register. The time is that of a cascaded carry-save network and not substantially greater than a single carry-propagate add time. This arises because all the adders are 'settling' (starting at the l.s. end), almost concurrently. There is approximately a 1-bit stagger between each stage. However, this is not consistent at all times due to the 7483 (quad full adder) that is widely used having an internal carry predictor. Thus, assuming the carry-in is settled (and the 9 inputs), the carry-out is true before the sum outputs. Furthermore, the latter are not generally produced with a straightforward stagger because of production spreads, etc. However, the time measured on several integrators indicated a worst case $\Delta Z/\Delta Y$ output delay of rather less than 1 μs. (Quite evidently, a factor of 8–10 improvement in the figures might be expected as a direct consequence of using a faster logic family, e.g. ECL.)

The total iteration time is thus 3 μs for the machine, as built. From this, the performance of the elements can be derived in a manner directly comparable with both g.p.d.c.s and analogue components.

8.2. *Slew Rate*

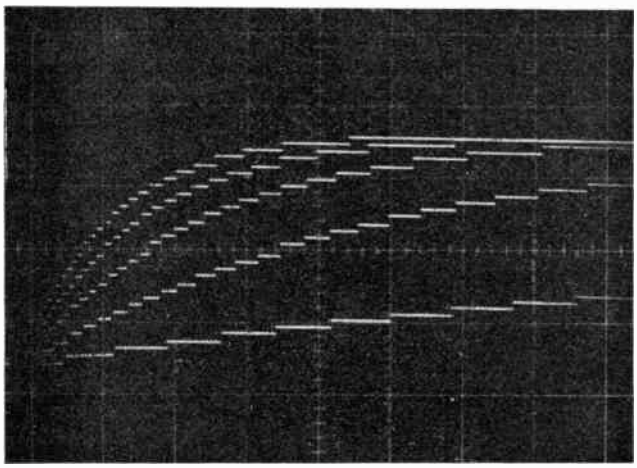
The trapezoidal algorithm, using 4-bit increments, and a 12-bit word gives a slew time from maximum negative to positive Y of

$$\frac{2^{12}\times 3}{16\times 10^{6}} = 770\ \mu s\ \text{(approx)}. \qquad \text{.....(8)}$$

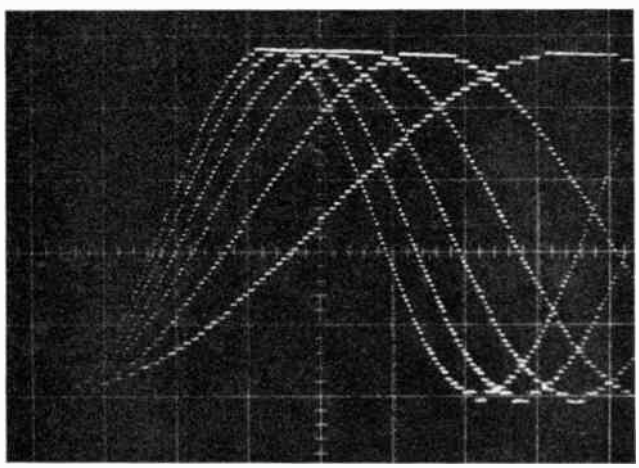
This is, of course, at the full 12-bit precision of the Y register and is not directly comparable with the slew rate of analogue comparators and the like. It is comparable with g.p.d.c.s.

8.3. *Sinewave Rate*

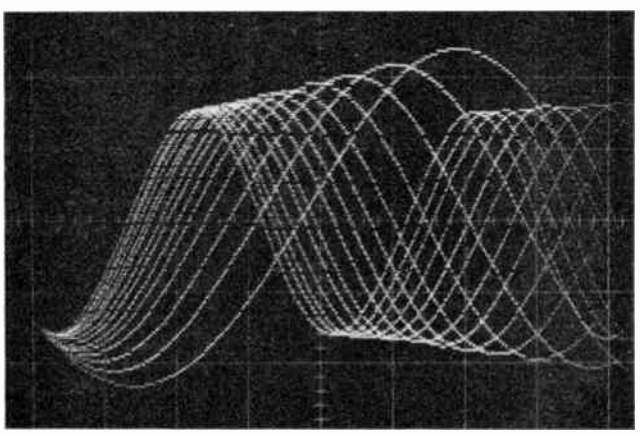
From this slew time the sinewave speed can be derived. Alternatively, it may be obtained by considering that if a sine/cosine loop is set up wherein a given integrator



(a) Solution of $\dot{y}+ky=0$, parameter k

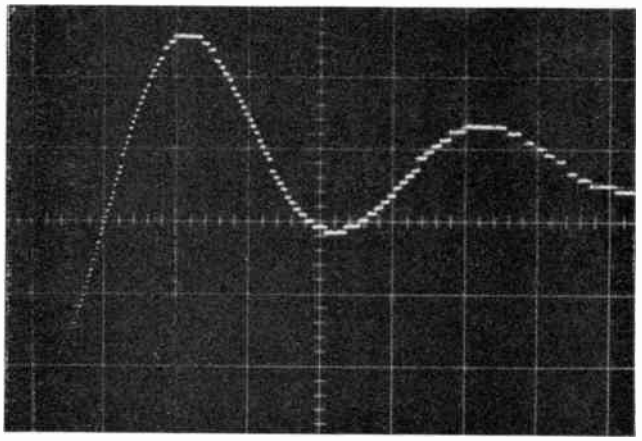


(b) Solution of $\ddot{y}+ky=0$, $\dot{y}_{(0)}\neq 0$, $y_{(0)}=0$, parameter k

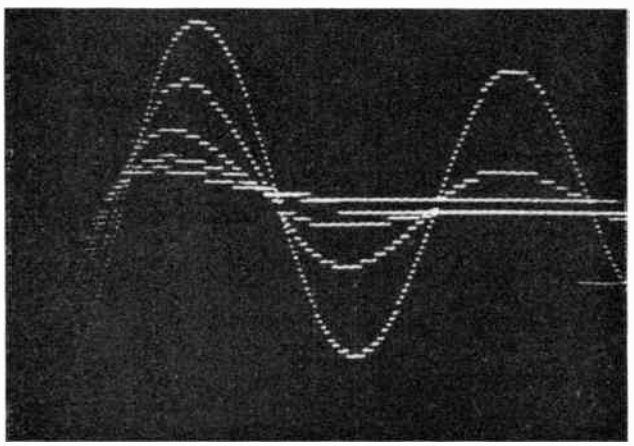


(c) Solution of $\ddot{y}+ky=0$, $\dot{y}_{(0)}\neq 0$, $y_{(0)}\neq 0$, parameter k

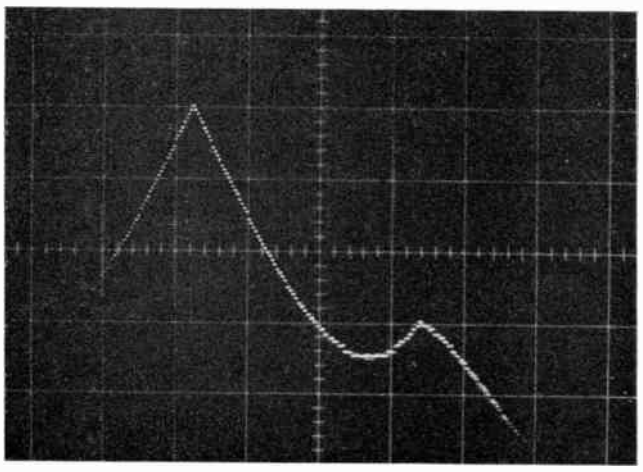
Fig. 9.



(a) Solution of $\ddot{y}+a\dot{y}+by=0$; $\ddot{y}$ displayed



(b) Solution of $\ddot{y}+k\dot{y}+ay=0$; $\ddot{y}$ displayed; $\ddot{y}$ parameter k



(c) Solution of third-order differential equation illustrating overloading of integrators

Fig. 10.

output ψ is given by:

$$\left.\begin{aligned}\psi &= A\sin(\omega t-\phi)\\ \dot{\psi} &= A\omega\cos(\omega t-\phi)\end{aligned}\right\} \qquad \text{.....(9)}$$

i.e. $\dot{\psi}_{(\phi)} = A\omega.$

The maximum slew rate (ξ) for ψ occurs at

$$t=\frac{2\pi K}{\omega}+\phi \quad \text{where} \quad K=0, 1, 2, \ldots$$

i.e. $\xi_{(\max)} = A\omega.$

Therefore, for this machine, the maximum sinewave speed is about 400 Hz. Due to the interaction between the truncation and round-off errors in the system, the amplitude remains constant for a large number of cycles and is certainly accurate to ± 1 bit of the Y register for 100 circles.

8.4 *Parameter Scanning*

The main advantages that have been obtained from the construction of this machine, apart from its solution, lie in its versatility for use in iterative computing situations. These include the ease and speed with which it may be loaded, reprogrammed, restarted or made to act on conditions arising during computation.

Compared with the conventional analogue computer, it can be loaded with both initial conditions and patching information directly from a paper tape reader, keyboard or computer interface. The loading speed is about 60 000 machine elements per second provided that the data source can match this rate. This applies whether the machine is being initially loaded or reprogrammed part way through a problem.

Iterative computing usually requires that problems be solved for a variety of initial conditions or potentiometer coefficients. It may even require changes to the equations 'patched'. In a hybrid computer, these changes are sometimes carried out by parallel logic specially patched in or by a digital computer. The program is usually held up for some time so that servo-set potentiometers, etc., may be actuated.

The d.d.a. can effect any of these changes very quickly just by changing the contents of one or more registers, and this is ideally suited for high-speed iterative computing. Consider, for example, the analysis of a network or structure which must be described by different equations at the boundaries. The d.d.a., without any physical modification, is capable of analysing the system, node by node, and reconfiguring itself as boundaries are approached.

Some examples of parameter scanning are shown in Figs. 9 and 10. These have been achieved through parallel logic which can be readily introduced to the system, as desired.

Figure 9(a) shows solutions to a simple first-order differential equation where the time-constant ($=1/k$) is varied as a parameter. As many values of k may be used as there are values of potentiometer fraction available. In this machine, P may be varied in steps of 1/128 from -1 to $+1$. Only five solutions have been shown in Fig. 9(a) for clarity. The waveform shown uses only the 6 most significant bits of the solution so that digitization is visible. Furthermore, as it is the content of a Y register that is displayed, small 'impulses' in the waveform are visible. These pre-empt a final change of value of Y, and indicate the action of the integrator logic implementing the trapezoidal algorithm. This causes the Y register to be offset by half the ΔY increment on any step.

Figure 9(b) shows a family of sinewaves being generated using various values of P to control the frequency. If non-zero conditions are used for both integrators, then varying P affects both the amplitude and frequency (Fig. 9(c)).

Figures 10(a) and (b) show solutions to second-order equations, once again, with a 6-bit digitization of the output. P registers can be used to vary both frequency and damping factor through a wide range of values.

Because 2's complement number notation is used, digital integrators do not saturate, but merely recycle through their range of values. Thus, if 1 is added to the maximum positive value that can be accommodated in a register, the maximum negative quantity results. In the solution of an equation, an overloading integrator will suffer a step change of value equal to its dynamic range (2^{12} in this machine). The integrator after it will then suffer a step change of slope as shown at the top of Fig. 10(c), and the next integrator a step change of slope rate.

8.5. *Loading*

The time to load a single integrator (with all 5 words of data) is limited in the present machine by the input device (a 200 characters per second paper tape reader). Even if a very fast parallel computer interface was provided, this would still be the case as the 8 MHz shift rate is capable of disposing of a byte (character) in 1 μs giving a load time inclusive of data assembly in the ΔZ registers of 14 μs. The 8 MHz rate, as has been stated, is a limitation imposed by the integrators' $\Sigma\Delta Y$ logic. If a separate clock were provided for data assembly/loading, the load time could be approximately halved as all active logic in this phase is only in the form of conventional shift registers.

8.6. *Accuracy of Computation*

The accuracy of the d.d.a. is governed in very much the same way as a g.p.d.c. Both data round-off effects and independent variable discretization (truncation) gradually erode the precision of solutions on a per step basis. The only major difference between the g.p.d.c. and the d.d.a. is that the former generally incurs round-off errors by permanently losing the lower bits of variables after multiplication by Δt or ΔX. The d.d.a., on a given step, suffers the same round-off, but the non-transmitted data are preserved in the R (residue) register to be accumulated with subsequent low significance integral increments. The round-off error therefore tends to manifest itself as a varying time or phase delay with a non-linear characteristic not unlike that exhibited by a g.p.d.c. algorithm.

The truncation errors are exactly akin to those of a g.p.d.c. algorithm, the effects of which may be determined by solution of the associated difference equations. The only point of divergence lies in the fact that both hardware and algorithmic approaches to integration inevitably allow the round-off and truncation errors to interact. Because the effects of the round-off error differ, this reflects, to a small extent, on the truncation errors that are apparently induced.

The machine built uses a trapezoidal integration algorithm to provide a first-order post-correction for the well known Euler method. This, for a wide variety of problems, shows good stability with both a predictable and moderately easily calculated error growth rate. Also, because of the high cost per integrator of implementing

the more exotic algorithms, this was, to some extent, an engineering compromise.

The machine built also employed a 4-bit transfer method as a compromise between the high round-off errors per step associated with, say, binary $\Delta(-1, +1)$ or ternary $\Delta(-1, 0, +1)$ transfer methods on the one hand, and transmission of the integral, *in toto*, on the other hand. The compromise on the latter part must necessarily depend on the expected number of iterations per solution for the problems likely to be encountered.

9. Machine Expansion

There are several ways in which this basic design can be extended both for the purposes of software simplification and reduction of the demands on outside hardware, particularly in the interface to the g.p.d.c.:

(a) To include in each integrator a separate 'Y' register which is loaded at the same time as the operational one. This would store the initial condition, i.e. would not be updated during computation, but would refresh the operational Y register, if necessary. This would allow more ready implementation of certain simulation languages, in particular, the Simulation Council Inc. version, CSSL.[11] In this, the updating of registers for each run is done on an exceptions-only basis, which is particularly powerful for fast parameter sweeping or hill-climbing. From the hardware viewpoint, it reduces the data flow through the interface and hence loading of the g.p.d.c.s I/O busbar. It does mean, of course, that selective or addressable updating must be built in rather than the hitherto simpler approach of a systematic refill scan through all the machine elements.

(b) As mentioned in the description of the interconnexions method, other machine topologies can be tried with a view to reducing the time taken for interconnexions. This is very important for large machines. The extension can also include the use of the space as well as time domain, i.e. by multiple ranking of selectors (channels). In the limiting case of use of the space domain only, a totally combinatorial system is achieved but is bound to be costly, and raises the problem of concurrently adding several input increments to any integrator. 'Treeing' of adders to do this, even with carry-save techniques, will accumulate many gate delays. A possible solution to this problem is the (partial) use of look-up tables. Such tables, at present generally manufactured as metal oxide silicon devices, are very fast but would need to be very capacious for high fan-in machine elements. A compromise solution may lay in using them just for summing each set of equally weighted bits from a group of incoming increments. A single device could then be time-shared for each significance of such increments.

As an example, a 256-word by 4-bit (1024 bits) m.o.s. read-only memory (r.o.m.) could sum 8 equally weighted bits to produce a 4-bit word of conventionally weighted bits. This, is not, at present, a standard chip design, but the introduction of programmable r.o.m.s (p.r.o.m.) has allowed non-standard contents to be electrically formed, thereby eliminating the high pattern design cost.

10. Conclusion

The d.d.a. built has indicated that, with technology and systems practices now available, it is justifiable to build this special form of hardware for real-time simulation. The speed advantage gained from special-purpose logic structures can more than offset the limited repertoire of the systems.

As little as five years ago it would have been, in most circumstances, very difficult to obtain a clear advantage from such techniques. Also, the development, in the last decade, of sophisticated design tools such as c.a.d., automatic circuit design/layout and logic fault detection methods, allows new electronic systems to come into being more quickly, and with a greater chance of initial success, when manufactured. From the system architecture viewpoint, the d.d.a. represents one of the structures most amenable to parallel processing techniques and this is, no doubt, due to the comparative simplicity of each machine element, in particular its interface with the outside world. This, together with the very small number of lines needed to effect this interface, make it an attractive logic base for large-scale integration.

11. Acknowledgments

The authors wish to acknowledge, with thanks, the support given to this project by Professor D. R. Chick of the Department of Electronic and Electrical Engineering, University of Surrey.

12. References

1. Owen, P. L., Partridge, M. F., and Sizer, T. R. H., 'CORSAIR —a digital differential analyser', *Electronic Engineering*, 32, p. 740, 1960.
idem, 'A transistor digital differential analyser', *J. Brit. I.R.E.*, 22, p. 83, August 1961.
2. Forbes, G. F., 'Digital Differential Analysers' (Pacoima, 1957).
3. Donan, J. F., 'The serial memory dda', *Computation*, 6, p. 102, 1952.
4. Von Handel, P. (Ed), 'Electronic Computers', pp. 139–209 (Springer Verlag, Vienna, 1961).
5. Benyon, P. R., 'A review of numerical methods for digital simulation', *Simulation*, November 1968, pp. 219–238.
6. Kopal, Z., 'Numerical Analysis' (Chapman & Hall, 1955).
7. Hyatt, G. P. and Ohlberg, G., 'Electrically alterable digital differential analyser', Proc. A.F.I.P.S. Spring Joint Computer Conference, 1968, pp. 161–169.
8. Sizer, T. R. H. (Ed.), 'The Digital Differential Analyser' (Chapman & Hall, London, 1968).
9. Bywater, R. E. H., 'One step integration method for digital differential analysers', *Electronics Letters*, 6, p. 613, 1970.
10. Hatvany, J., 'The d.d.a. integrator as the iterative module of a variable structure process control computer', *Automatica*, 5, No. 1, pp. 41–9, 1969.
11. Strauss, Jon C. (Ed.), 'The SCi continuous system simulation language (CSSL)', *Simulation*, December 1967, pp. 281–303.

Manuscript received by the Institution on 29th November 1971.
(Paper No. 1446/Comp. 140.)

U.D.C. 621.372.833

Three-resonant-mode Adjustment of the Waveguide Circulator

J. HELSZAJN,
M.S.E.E., Ph.D., C.Eng., F.I.E.R.E.†

SUMMARY

In its matched condition the scattering matrix eigenvalues of the three-port circulator lie equally spaced on a unit circle. As the frequency is varied these eigenvalues rotate in the complex plane at different speeds and the phase relation between them no longer corresponds to an ideal circulator. This is primarily due to the fact that the reference eigenvalue is normally associated with a non-resonant field pattern and the other two are associated with resonant ones. The purpose of this paper is to describe one circulation adjustment in which all three eigenvalues are associated with resonant field patterns. Experimental results obtained on a waveguide circulator are given for these two different adjustment procedures.

† *Senior Research Fellow, Department of Electrical and Electronic Engineering, Heriot-Watt University, Edinburgh EH1 1HX.*

1. Introduction

In its matched condition the three eigenvalues of the scattering matrix of the 3-port junction circulator lie equally spaced on a unit circle in the complex plane. To obtain this eigenvalue arrangement it is necessary to adjust the phases of two of them with respect to the third one, and so at least two independent physical variables are required.[1,2] These are usually taken as the diameter of the ferrite disk and the magnitude of the direct magnetic field. In the unmagnetized configuration two out of the three eigenvalues are degenerate. In this adjustment procedure the non-degenerate eigenvalue is associated with a non-resonant field pattern and the pair of degenerate eigenvalues are associated with resonant ones. The eigenvalues are reflexion coefficients associated with possible different ways of exciting the junction.

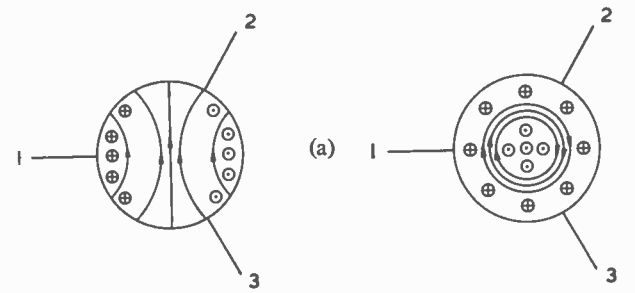
It is also possible to construct junction circulators in which all three eigenvalues are associated with resonant field patterns. This new adjustment procedure has recently been described independently.[3,5] If the Q-factors of the three field patterns are made equal then the three eigenvalues will all have the same frequency variation. The bandwidth of the circulator will then be very wide.

The purpose of this paper is to present the experimental adjustment of such a waveguide circulator. This is done by introducing one additional independent variable in the form of a thin metal pin through the centre of the ferrite disk. This additional variable can rotate the reference eigenvalue by 180° in the complex plane. It can therefore be used to establish an additional resonant field pattern within the ferrite post.[1,6,7,8] In this way all the eigenvalues are associated with resonant field patterns. The paper includes experimental results obtained on a waveguide circulator for the two different adjustment procedures described.

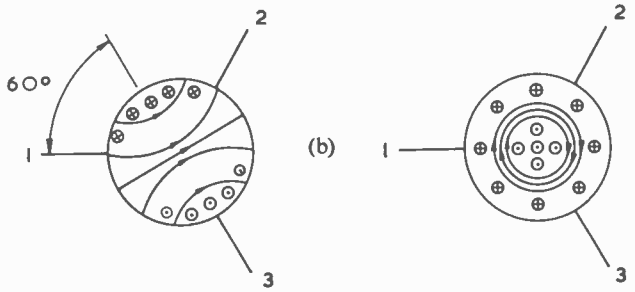
The possibility of widebanding waveguide junction circulators using dielectric sleeves and metal posts extending the full height of a waveguide has been predicted from a computer-aided design.[9,13] In this approach the additional field pattern is established within the overall junction rather than in the region of the ferrite post. This is because a full height post presents an inductive reactance at the centre of the junction and a partial height post presents a capacitive one. The two configurations therefore employ different field distributions with corresponding sets of scattering coefficients in each case.

2. Circulation Adjustment of Three-resonant-mode Circulator

One method of adjusting symmetrical microwave junctions consists of adjusting the eigenvalues of the scattering matrix by symmetric perturbations of the junction until the required eigenvalue arrangement is satisfied. Each such perturbation of the junction leads to a different scattering matrix whose entries can be calculated by taking linear combinations of the eigenvalues. The eigenvalues of the scattering matrix are the reflexion coefficients associated with the different ways of exciting the junction. For a lossless junction these reflexion coefficients have unit amplitude because they



(a) $n = \pm 1$ and $n = 0$ field patterns in unmagnetized ferrite disk.



(b) $n = \pm 1$ and $n = 0$ field patterns in magnetized ferrite disk.

Fig. 1.

are associated with reactive resonant or non-resonant networks. The physical perturbation of the junction is simplified if the electromagnetic field patterns of the required resonant networks are known. This method has been described in detail elsewhere.[6,8] In principle it is possible to construct junction circulators in terms of the reflexion coefficient only.

In the conventional adjustment of the 3-port junction using only two resonant modes the scattering matrices which must be established have the following entries for the reflexion coefficients

$$|S_{11}| = \tfrac{1}{3} \quad \ldots.(1)$$

$$|S_{11}| = 0 \quad \ldots.(2)$$

The first reflexion coefficient is obtained in the usual way by adjusting the diameter of the ferrite disk. This establishes the $n = \pm 1$ modes within the junction. The second reflexion coefficient is obtained by magnetizing the junction. This independent variable rotates the standing wave formed by the $n = \pm 1$ modes through 30° thereby placing a null of the standing wave at port 3. The field patterns for this arrangement have been described elsewhere.[14,15]

In the case of the three-resonant-mode circulator described in this paper the circulation condition is satisfied by obtaining the following three reflexion coefficients.

$$|S_{11}| = \tfrac{1}{3} \quad \ldots.(3)$$

$$|S_{11}| = 1 \quad \ldots.(4)$$

$$|S_{11}| = 0 \quad \ldots.(5)$$

The first reflexion coefficient is again obtained by tuning the $n = \pm 1$ modes within the junction by varying the diameter of the ferrite disk. The second reflexion coefficient is obtained by introducing a thin metal post through the centre of the junction. This independent variable tunes the $n = 0$ mode within the junction. The last reflexion coefficient is obtained by rotating the standing wave formed by the $n = \pm 1$ modes through 60° by magnetizing the junction. The detailed eigenvalue diagrams leading to the above reflexion coefficients have been described in another paper.[5]

Figure 1(a) shows the $n = \pm 1$ and $n = 0$ modes in an unmagnetized disk. Figure 1(b) shows the $n = \pm 1$ and $n = 0$ modes in a magnetized disk in which the former modes are rotated by 60°. If the amplitudes of the electric fields of the two modes are the same at the edge of the disk they will cancel at port 3 and transmission will take place between ports 1 and 2.

3. Experimental Circulator

The experimental work was carried out on a waveguide junction. The configuration used is shown in Fig. 2. The material used was a garnet with saturation magnetization of 0·175 Wb m^{-2}. The relative dielectric constant of the garnet material was $\varepsilon_r = 14{\cdot}5$. The waveguide size used was WR 187 and the operating frequency was about 5·5 GHz. The garnet disk was 12·2 mm in diameter with a 2·5 mm hole through its centre to take the tuning post.

The trimming of the tuning post was done with a micrometer. Spring fingers were used to make contact to the tuning post where the latter entered the waveguide. The direct magnetic field was applied using a coil.

4. Experimental Adjustment of Circulator Using Two Resonant Modes

The conventional adjustment of junction circulators has been described in Section 2. In this adjustment the two usual variables are the diameter of the ferrite disk and the magnitude of the direct magnetic field or the magnetization of the ferrite material.

Using this approach, the first circulation condition is obtained by determining the frequency at which the reflexion coefficient is a minimum. For an ideal circulator equation (1) shows that this condition coincides with $S_{11} = \frac{1}{3}$. Figure 3 shows the variation of the return loss as a function of frequency. From the illustration, the operating frequency is given as 5·47 GHz which coincides with the maximum return loss of 9·5 dB. The second circulation condition is now obtained simply by adjusting the magnitude of the direct magnetic field until $S_{11} = 0$. Figure 4 shows the variation of the return loss as a function of direct magnetic field. An ideal circulator is obtained at $H_0 = 6{\cdot}7$ kA m^{-1}.

The overall bandwidth of the device is shown in Fig. 5. The bandwidth obtained here is 2·9%, which is typical for a directly coupled junction of this type.

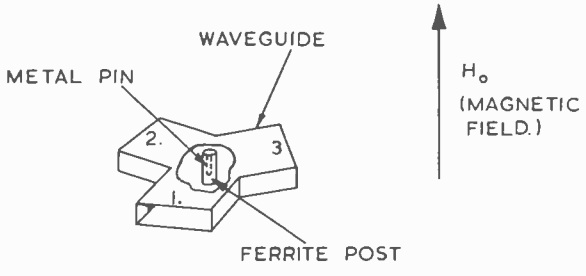


Fig. 2. Configuration of experimental waveguide circulator.

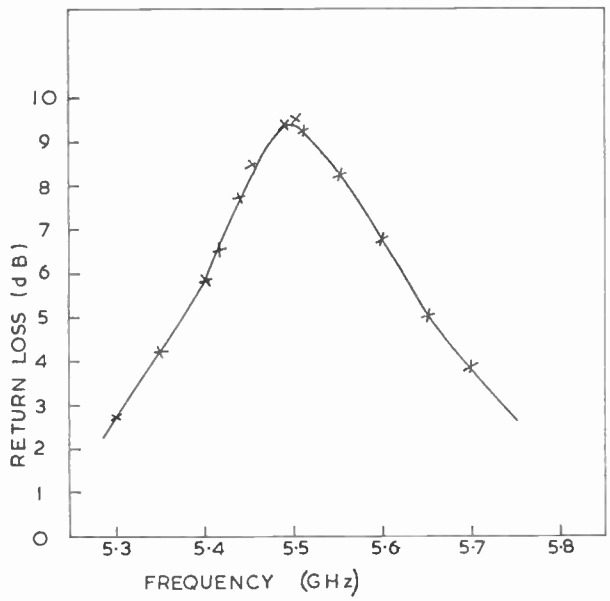


Fig. 3. Variation of return loss with frequency for 12·2mm diameter garnet disk.

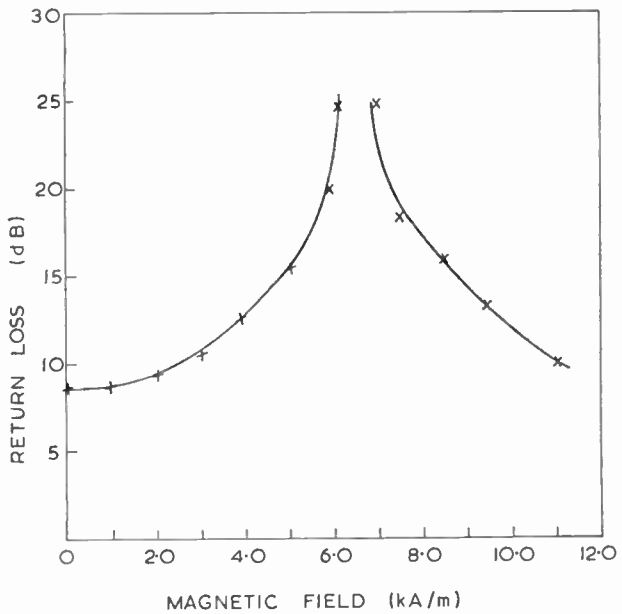


Fig. 4. Variation of return loss with direct magnetic field for 12·2 mm diameter garnet at 5·47 GHz.

5. Experimental Adjustment of Circulator Using Three Resonant Modes

The experimental adjustment of junction circulators using three resonant modes instead of two can be described by making use of the arrangement shown in Fig. 2. The first adjustment is common to both possible circulation adjustments and so Fig. 3 satisfies the first perturbation of the junction here also. This adjustment establishes the $n = \pm 1$ field patterns within the junction. The second circulation adjustment is now obtained by introducing one additional resonant field pattern within the junction. One suitable field pattern is the $n = 0$ one which can be tuned by introducing a thin metal post through the centre of the junction. This perturbation of the junction is satisfied when S_{11} passes through unity. The variation of the return loss, as the length of the tuning post is varied, shown in Fig. 6, indicates that the return loss passes through a minimum when the length of the metal post is 5·5 mm. This perturbation of the junction therefore establishes the $n = 0$ field pattern within the junction. This post has the same electrical length as the one used to establish the $n = 0$ mode in the construction of the single junction 4-port circulator.[7] The final perturbation of the junction is now obtained by applying a static magnetic field to the device. Figure 7 shows the variations of the return loss as a function of the direct magnetic field. An ideal circulator is obtained with $H_0 = 8{\cdot}5\ \text{kA m}^{-1}$. This last perturbation of the junction

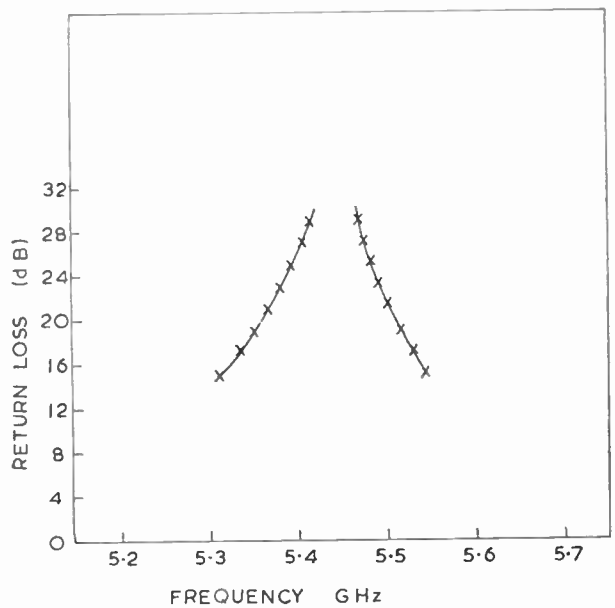


Fig. 5. Variation of return loss with frequency for two-resonant mode circulator.

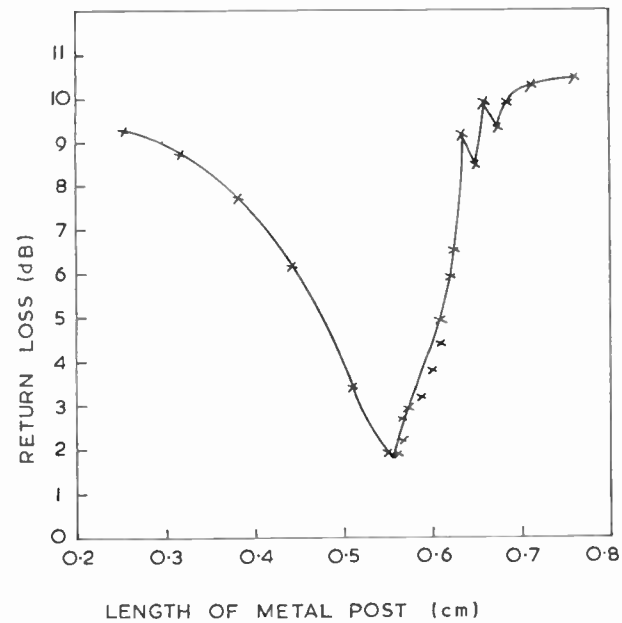


Fig. 6. Variation of return loss with length of metal post for 12·2 mm diameter garnet disk at 5·47 GHz.

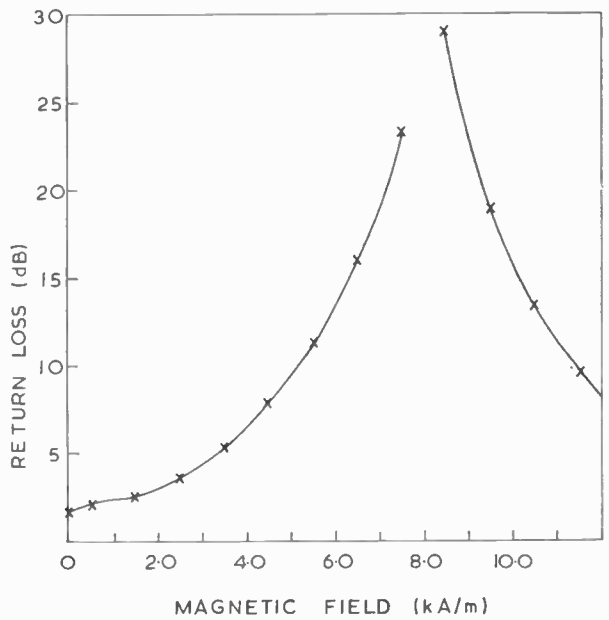


Fig. 7. Variation of return loss with direct magnetic field for 12·2 mm diameter garnet post at 5·47 GHz with 5·5 mm long metal post.

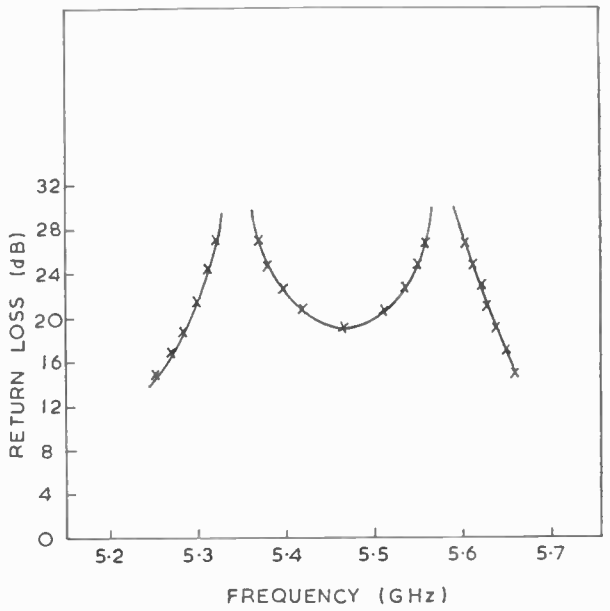


Fig. 8. Variation of return loss with frequency for 3 resonant mode circulator.

yields the reflexion coefficient of an ideal circulator.

Figure 8 shows the variation of the return loss as a function of the frequency. The bandwidth obtained here is about 7%. In obtaining this illustration a slight perturbation of the length of the metal post from about 5·5 mm to 5·3 mm was employed. A similar response has been obtained in the case of the lumped element circulator.[3,4]

6. Conclusion

This paper has described the experimental adjustment of the three-resonant-mode waveguide circulator. This adjustment requires three independent physical variables instead of the two necessary for the more usual two-resonant-mode junction. The three physical variables used in the case of the waveguide circulator are the diameter of the ferrite post, the length of a thin metal post through the centre of the ferrite post, and the magnitude of the direct magnetic field.

No attempt has been made so far to adjust the Q-factors of the three modes in order to obtain very wideband circulation.

7. Acknowledgments

The author wishes to thank Mr. William Shearer for making the measurements and Microwave and Electronic Systems Limited, Lochend Industrial Estate, Newbridge, Midlothian, Scotland, for providing the experimental facilities.

8. References

1. Auld, B. A., 'The synthesis of symmetrical waveguide circulators' *Trans. I.R.E. on Microwave Theory and Techniques*, MTT-7, pp. 238–46, April 1959.
2. Milano, U., Saunders, J. H. and Davis, L., Jnr., 'A Y-junction stripline circulator', *I.R.E. Trans.*, MTT-8, pp. 346–51, May 1960.
3. Konishi, Y., 'A high power u.h.f. circulator', *I.E.E.E. Trans.*, MTT-15, pp. 700–8, December 1967.
4. Knerr, R. H., Barnes, C. E. and Bosch, F., 'A compact broadband thin-film lumped element L-band circulator', *I.E.E.E. Trans.*, MTT-18, pp. 1100–8, December 1970.
5. Helszajn, J., 'Wideband circulator adjustment using $n = \pm 1$ and $n = 0$ electromagnetic-field patterns', *Electronics Letters*, 6, pp. 729–31, 12th November 1970.
6. Montgomery, C. G., Dicke, R. H. and Purcell, E. M., 'Principles of Microwave Circuits' (McGraw-Hill, New York, 1948).
7. Helszajn, J. and Buffler, C. R., 'Adjustment of the 4-port single junction circulator', *The Radio and Electronic Engineer*, 35, pp. 357–60, June 1968.
8. Helszajn, J., 'The adjustment of the m-port single junction circulator', *I.E.E.E. Trans.*, MTT-18, pp. 705–11, October 1970.
9. Davies, J. B., 'An analysis of the m-port symmetrical H-plane waveguide junction with central ferrite post', *I.R.E. Trans.*, MTT-10, pp. 596–604, November 1962.
10. Davies, J. B., 'Theoretical design of wideband waveguide circulators', *Electronics Letters*, 1, pp. 60–61, May 1965.
11. Parsonson, C. G., Longley, S. R. and Davies, J. B., 'The theoretical design of broadband 3-port waveguide circulators', *I.E.E.E. Trans.*, MTT-16, pp. 256–8, April 1968. (Letters).
12. Davis, L. E., 'Central-pin tolerances in broadband 3-port waveguide circulators', *Electronics Letters*, 4, pp. 307–9, July 1968.
13. Castillo, J. B., Jnr. and Davis, L. E., 'Computer-aided design of 3-port waveguide junction circulators', *I.E.E.E. Trans.*, MTT-18, pp. 25–34, January 1970.
14. Fay, C. E. and Comstock, R. L., 'Operation of the ferrite junction circulator', *I.E.E.E. Trans.*, MTT-13, pp. 15–27, January 1965.
15. Helszajn, J., 'A ferrite ring stripline junction circulator', *The Radio and Electronic Engineer*, 32, pp. 55–60, July 1966.

Manuscript first received by the Institution on 12th November 1971 and in final form on 10th January 1972. (Paper No. 1447/CC127).

U.D.C. 557.510.5: 621.317.335.3

An Improved Radio-frequency Probe for the Measurement of Electron Densities in the Ionosphere

H. W. BRYAN,*

J. WALL,*

and

J. H. WAGER*

SUMMARY

This paper describes a radio frequency technique for measuring ionospheric electron densities. Extensive use is made of linear integrated circuits and their application is discussed in some detail. Two alternative forms of the instrument are described, one measuring electron number densities from 10^2 to 10^6 cm^{-3} in four linear ranges, the other covering the same variation in two overlapping logarithmic response ranges.

* *Department of Space Research, University of Birmingham. Birmingham B15 2TT.*

List of Symbols

C capacitance between plates in the presence of free electrons

C_0 capacitance between plates in the absence of free electrons

C_s stray capacitance of support tube (Fig. 3)

e electronic charge

f probing frequency (hertz) $= \omega/2\pi$

i_0 r.f. current through sensor capacitance, C_0 (Fig. 3)

i_s r.f. current through stray capacitance C_s (Fig. 3)

K Boltzmann's constant

m electron mass

N electron density (m^{-3})

T absolute temperature degK

ε permittivity in the presence of electrons

ε_0 permittivity in free space

1. Introduction

Measurements of electron densities in the ionosphere from rockets and satellites have been made for a number of years by the space research group at Birmingham University using a radio frequency probe technique.[1-4] During this time the instrumentation has undergone considerable changes in order to take advantage of the advances in components technology. This present paper describes in detail the instrumentation currently being used which makes extensive use of linear integrated circuits.

Two versions of the instrumentation are discussed, each designed to measure electron densities in the range 10^2–10^6 electrons cm^{-3}, one using four linear output channels, the other two logarithmic outputs to cover the range of densities. The logarithmic channels retain their calibration accuracy over the full operating temperature range of 0–50°C without recourse to temperature-controlled logarithmic elements.

2. Theory of the Method

The operation of the radio-frequency electron density probe is based upon measurements of the permittivity of the plasma medium which contains an equal population of free electrons and positive ions, the measurement being made in terms of the change in capacitance of a parallel plate capacitor.

It has been shown[5, 6] that the permittivity of a plasma medium is a function both of the free electron content and of the probing frequency used. At the altitudes at which measurements are made certain simplifying assumptions may be applied, since the probing frequency is chosen to be high compared with both the electron collision frequency and the gyro-frequency. This gives rise to the following relation between permittivity, probing frequency and electron number density.

Using m.k.s. units,

$$\varepsilon = \varepsilon_0 \left(1 - \frac{Ne^2}{\varepsilon_0 m\omega^2}\right) \quad \text{F/m} \qquad \ldots\ldots(1)$$

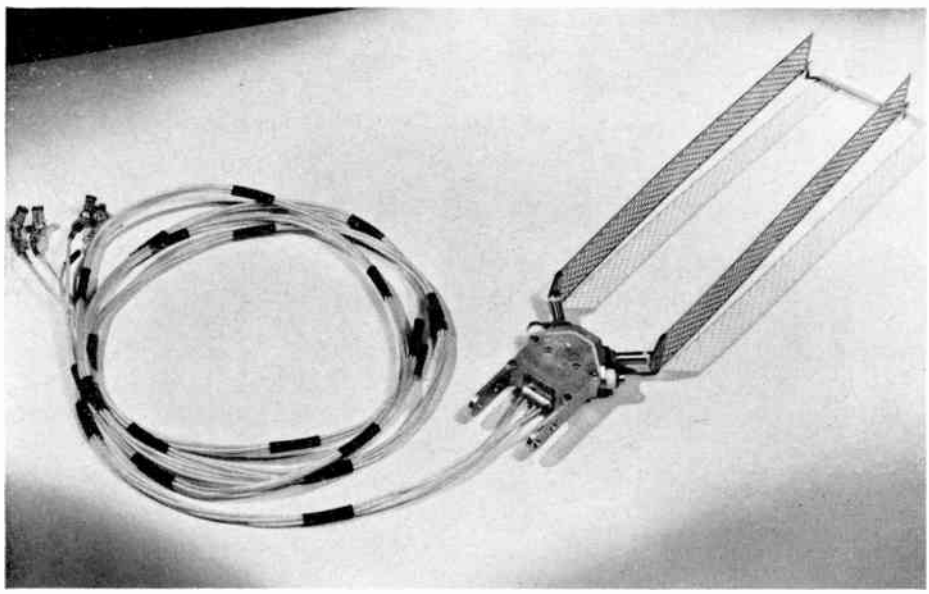


Fig. 1. The sensor.

From this relationship it may be seen that if the probing frequency is known the changes in permittivity observed are directly proportional to the electron density.

2.1. *The R.F. Probe: Principle of Operation*

The measurement of permittivity required from equation (1) may be made by observing the changes in r.f. current flowing in a parallel plate capacitor excited at a known fixed radio frequency. It is convenient to rewrite equation (1) in the following form:

$$C = C_0\left(1 - \frac{80{\cdot}6N}{f^2}\right) \text{ F} \qquad \text{.....(2)}$$

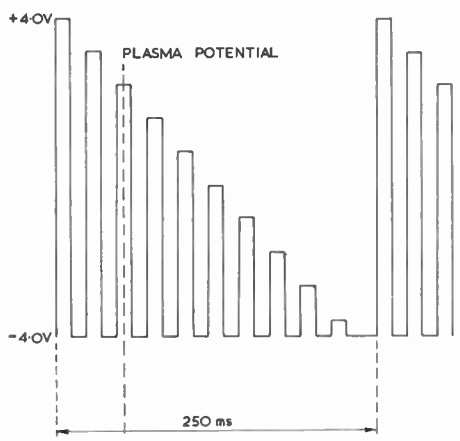
The frequency of operation of the probe described here is 39 MHz, hence equation (2) becomes

$$1 - \frac{C}{C_0} = 5{\cdot}3 \times 10^{-14}N \qquad \text{.....(3)}$$

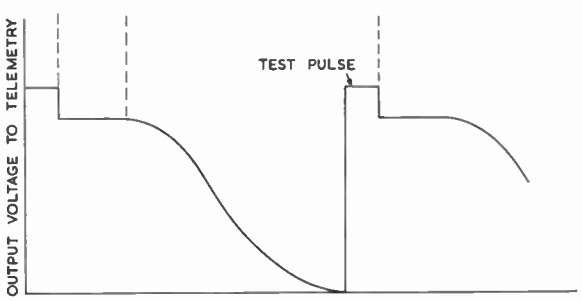
The capacitor probe takes the form of two grids each approximately 25 cm long by 2·5 cm wide and placed 6·4 cm apart as shown in Fig. 1. This probe is carried at the tip of a boom structure deployed from the space vehicle. If the probe is biased so that it is at the same potential as the plasma the volume between the grids will be filled with electrons at the ambient density giving a value of C for the capacitance of the sensor. If the probe is biased negatively with respect to the plasma potential electrons will be excluded from the volume between the grids giving a value of C_0 for the capacitance of the sensor. Thus a variation in the d.c. bias voltage applied to the sensor structure will cause changes in the magnitude of the r.f. current flowing between the grids.

Since the r.f. currents flowing between the grids of the sensor may be varied by means of a d.c. bias voltage applied to the structure it follows that if the bias changes are carried out by a square-wave whose negative level is −4·0 V and whose positive level is plasma potential the r.f. currents flowing in the sensing capacitance will be amplitude modulated by a square wave. In practice the square-wave frequency used for the bias changes is 6 kHz.

However, the exact value of plasma potential is dependent on a number of factors and it is necessary to use a square wave whose maximum positive amplitude



(a) Chopped ramp to sensor.



(b) Data signal.

Fig. 2.

gradually decreases with time as shown in Fig. 2(a) on which a typical value for plasma potential is marked. The behaviour at potentials more positive than plasma potential is discussed more fully elsewhere.[7]

2.2. *Sensor Design*

A detailed diagram of the sensor is shown in Fig. 3 and its operation is as follows.

The two sensor grids are carried on short support tubes and are connected directly to either end of a centre tapped coil which is tuned by the sensor, trimmer and stray capacitances (amounting to $(C_1 C_2/C_1+C_2)+(C_s/2)+C_0$) to the probing frequency of 39 MHz. Surrounding each support tube is a ferrite ring. These are linked by a single turn loop which feeds a signal to the r.f. amplifier. This arrangement ensures that this signal is due only to the r.f. currents flowing through C_0 since the currents i_s through the stray capacitances of the support tube do not link the ferrite and therefore do not induce a signal into the pick up loop.

The centre tap of the coil is connected to the grid support tubes and to this point is applied the chopped ramp signal shown in Fig. 2(a). Thus the grids and their supports are at the same d.c. potential, but the support tubes carry no r.f. signal by virtue of the circuit symmetry and the capacitance of the cable feeding the chopped ramp from the electronics package.

It has been shown in the previous section that the magnitude of the r.f. current flowing between the grids, i.e. through C_0, is proportional to the ambient electron density. Thus, as the d.c. potential on the grids is chopped the electrons are alternately removed from and introduced into C_0 and the signal level fed to the r.f. amplifier changes at the chopping rate. The r.f. amplifier input is therefore an amplitude modulated r.f. signal, the modulation envelope of which is a 6 kHz square-wave. The maximum depth of modulation is a function of electron density.

3. The Block Diagram

Referring to Fig. 4, the sensor is driven by the r.f. output from a crystal-controlled oscillator at 39 MHz, the amplitude of r.f. appearing between the sensor grids being controlled at 3·0 V r.m.s. by means of a feedback loop.

The signal from the pick-up loop on the sensor is amplified and demodulated to yield a 6 kHz signal whose amplitude is a function of the ambient electron density.

The output of this demodulator feeds either the linear or the logarithmic output stages depending upon which version of the equipment is being used. The linear output stages consist of four a.c. amplifiers of differing gains each feeding a synchronous detector to give four output channels (ED1–ED4) which, between them cover the whole measurement range of electron density.

The logarithmic output stages consist of two linear a.c. amplifiers of differing gain feeding synchronous detectors followed by the logarithmic output stages and are described in detail in Section 4.

Close attention to component selection has been necessary in the design of the synchronous detectors, in order to achieve linearity over the three decades of input handled by the output stages. Particular problems involved are: achieving low saturation resistance in the chopping device and low coupling between reference input and output terminals, and selection of the integrating amplifiers for low temperature coefficient of input bias current.

Since the changes in sensor capacitance with free electron density are small (only 5·3% change for an electron density of 10^6 electrons per cm^3) the r.f. amplifier and oscillator are fed from a separate, well stabilized supply to avoid spurious modulation.

The chopped ramp voltage which is applied to the sensor and illustrated in Fig. 2(a) is generated by using

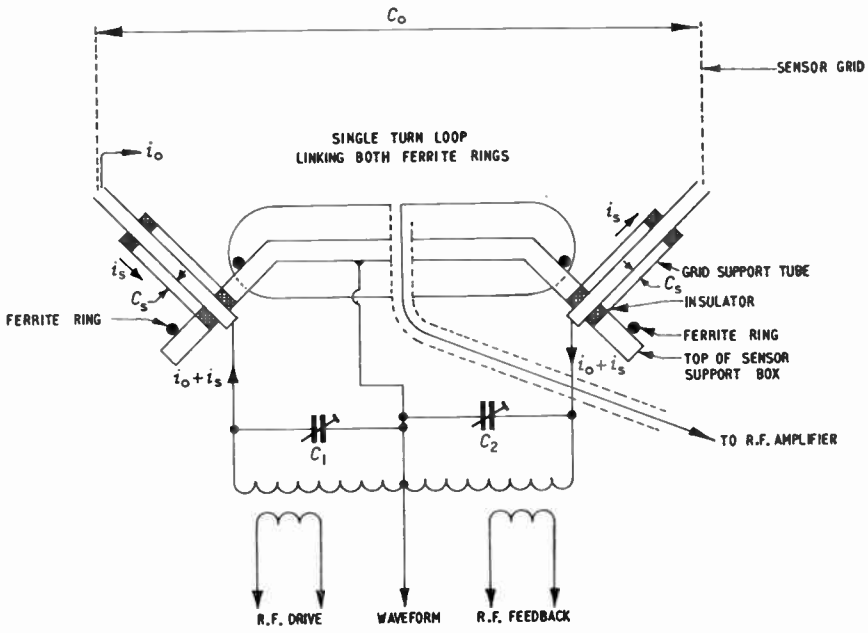


Fig. 3. Functional diagram of sensor.

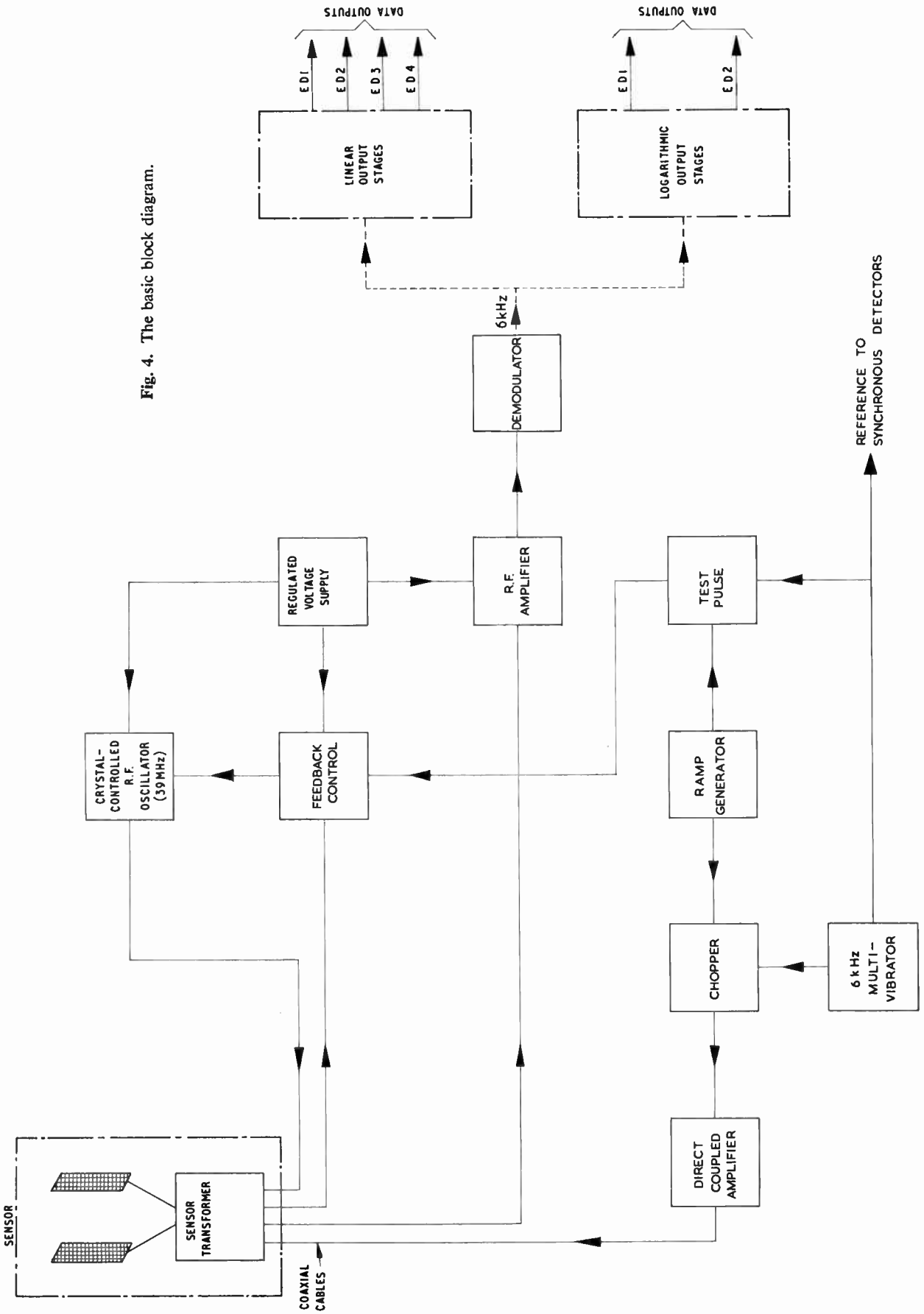


Fig. 4. The basic block diagram.

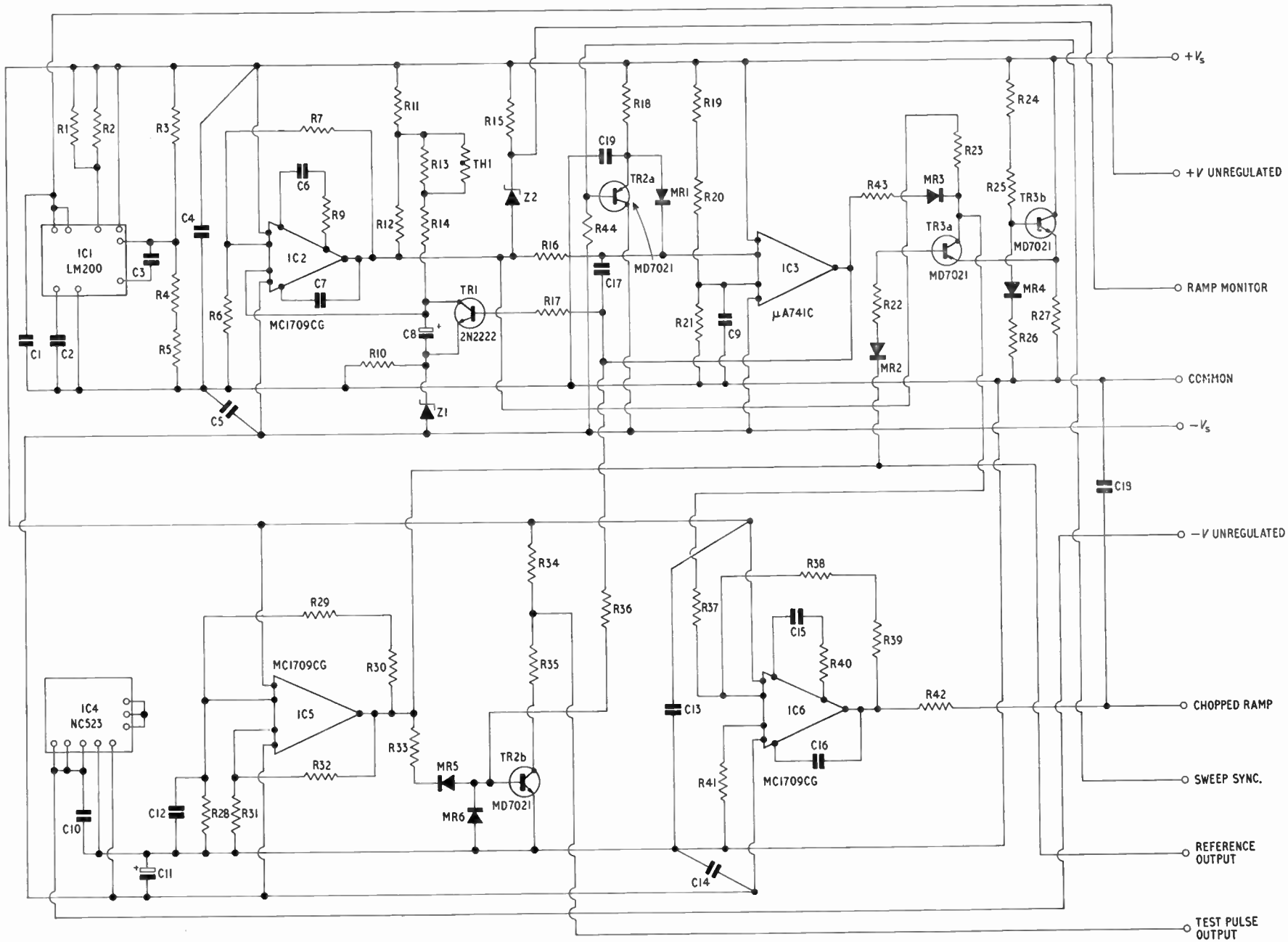


Fig. 5. The chopped ramp generator circuit.

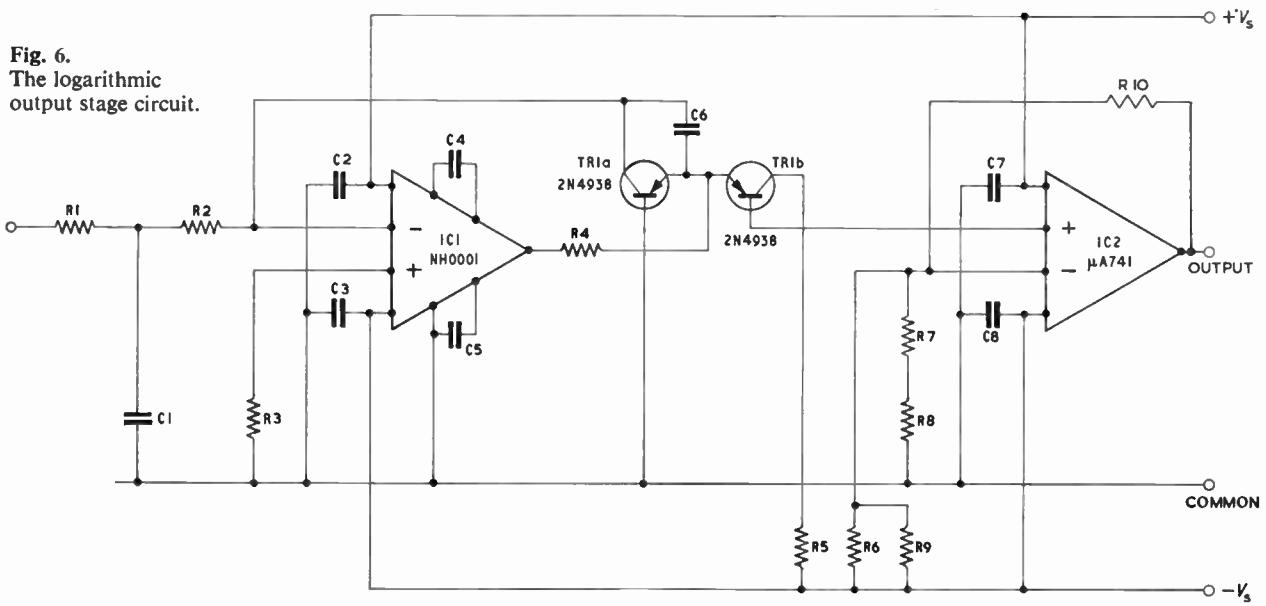


Fig. 6.
The logarithmic output stage circuit.

a ramp generator to control the reference level of a chopper operating at 6 kHz. The output of the chopper is taken to a low-gain, directly-coupled amplifier whose gain, bandwidth and offset are adjusted to give the output levels shown in Fig. 2(a).

From Sections 2.1 and 2.2 it will be seen that modulation of the r.f. and hence the generation of a data signal is dependent upon free plasma electrons being present. To facilitate ground checkout procedures (when of course there are no free electrons) it is useful to have some form of signal on the data channels. A short burst of 6 kHz modulation is therefore applied to the r.f. oscillator to form a test pulse which precedes the normal data signal obtained in space, as shown in Fig. 2(b).

The reduction of the data signal to an electron density value requires only a measurement of the amplitude of the flat portion of the curve and translation to a value of C/C_0 via a calibration factor for the instrument.

4. Circuit Details

The majority of the circuit functions are performed by linear integrated circuit operational amplifiers and the circuits in general require little comment since they are relatively conventional. The chopped ramp generator and logarithmic amplifiers however are rather more specialized and will be dealt with in some detail.

4.1. *The Chopped Ramp Generator*

The circuit is shown in Fig. 5. A 250 ms ramp is generated by IC2, TR1 and IC3 using a 'boot-strap' constant current circuit with its output taken from IC2 via R23 to the chopper TR3a which is driven from the 6 kHz multivibrator formed by IC5.

The chopper output is fed to amplifier IC6 which is directly coupled and of adequate bandwidth and gain to give an output waveform as shown in Fig. 2(a).

The need for a test pulse was discussed earlier in the text and the burst of 6 kHz to modulate the r.f. oscillator is generated by gating the multivibrator output via MR5 with a square wave derived from the ramp generator flyback circuit. Transistor TR2b merely provides 180° of phase shift in order that the test pulse appearing on the data signal shall be of the correct polarity.

Some ionospheric experiments require that the ramp shall be synchronized to other equipment and this is achieved by applying a sync signal via transistor TR2a to initiate the ramp flyback.

4.2. *The Logarithmic Amplifier*

The circuit of one of the two logarithmic output stages is shown in Fig. 6. The logarithmic function is generated by utilizing the fact that the collector current (I_c) of a bipolar transistor is proportional to the logarithm of the base-emitter voltage (V_{be}).[8]

Thus:

$$I_c \propto \exp\left(\frac{eV_{be}}{kT}\right)$$

from which it follows that

$$V_{be} \propto \frac{KT \ln I_c}{e} \qquad \text{.....(4)}$$

This relationship holds true over a much wider range of currents than the more commonly used V/I characteristic of a semiconductor diode. If two identical transistors are used as the logarithmic element in such a way that their base-emitter voltages are opposing, the relationship between them is

$$\Delta V_{be} = \frac{KT}{e} \ln \frac{I_{c_1}}{I_{c_2}} \qquad \text{.....(5)}$$

where $\Delta V_{be} = V_{be_1} - V_{be_2}$.

Referring to Fig. 6, the matched transistor pair is TR1a and TR1b. TR1a is connected with its base emitter junction across the output of operational amplifier IC1, the negative feedback for this amplifier

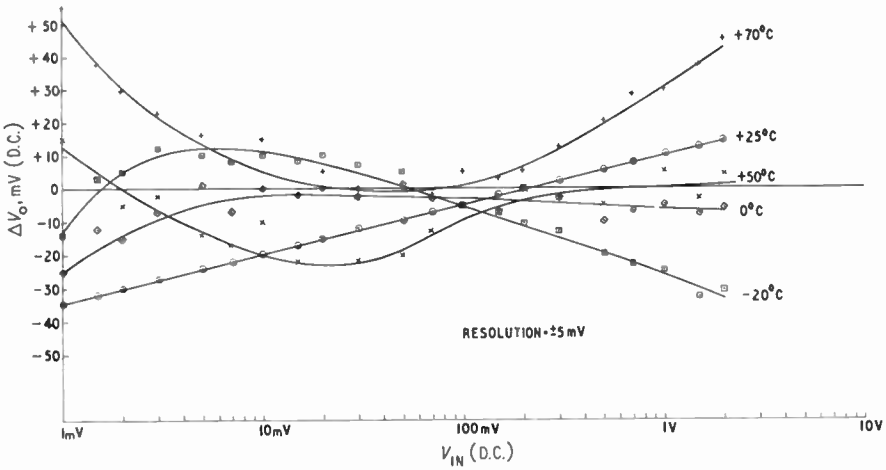


Fig. 7. Output/input voltage error for varying temperatures.

being provided by the collector current of the transistor. The second transistor of the matched pair is emitter coupled to the first and provided with a constant collector current given by

$$I_{c_2} = \frac{-V_s}{R_5}$$

where $-V_s$ is the negative supply voltage.

The voltage appearing at the base of TR1b is thus proportional to the logarithm of the collector current of the first transistor. Since this collector forms the feedback path for the virtual earth operational amplifier IC1, the current through it must equal the input current which can be expressed as

$$I_{in} = \frac{E_{in}}{R_{in}}$$

Substituting in equation (5) gives:

$$\Delta V_{be} = \frac{KT}{e} \ln \frac{E_{in}}{R_{in}} \times \frac{R_5}{-V_s}$$

If ΔV_{be} is then amplified by a further operational amplifier IC2 having a gain equal to M_2 the overall transfer function is

$$V_0 = M_2 \frac{KT}{e} \ln E_{in} \times \frac{R_5}{-V_s \times R_{in}} \qquad \ldots\ldots(6)$$

where V_0 is the output voltage of IC2.

If the forward gain of IC2 is made temperature-dependent by using a temperature-sensitive element in the feedback network, the term $M_2 . KT/e$ in equation (5) may be made a constant leaving

$$V_0 = K_1 \ln K_2 E_{in} \qquad \ldots\ldots(7)$$

The published forms of this circuit[8, 9] use a dual n–p–n transistor as the logarithmic element since these tend to have a good performance over a wide dynamic range. However, the present system required that the output signal should be positive, should increase positively with increasing input signal and should have an offset adjustment so that the zero crossover could be selected as required. These features together meant that a p–n–p transistor had to be used since the output stage (IC2) had to be non-inverting in order that its feedback current did not adversely affect the operation of the logarithmic element. Careful choice of the type of p–n–p transistor used has resulted in a very good matching of the base-emitter characteristics over a wide range of temperatures, collector leakage currents causing only negligible errors in the lower of the three decades used.

Referring again to Fig. 6, the residual a.c. component in the negative going signal from the synchronous

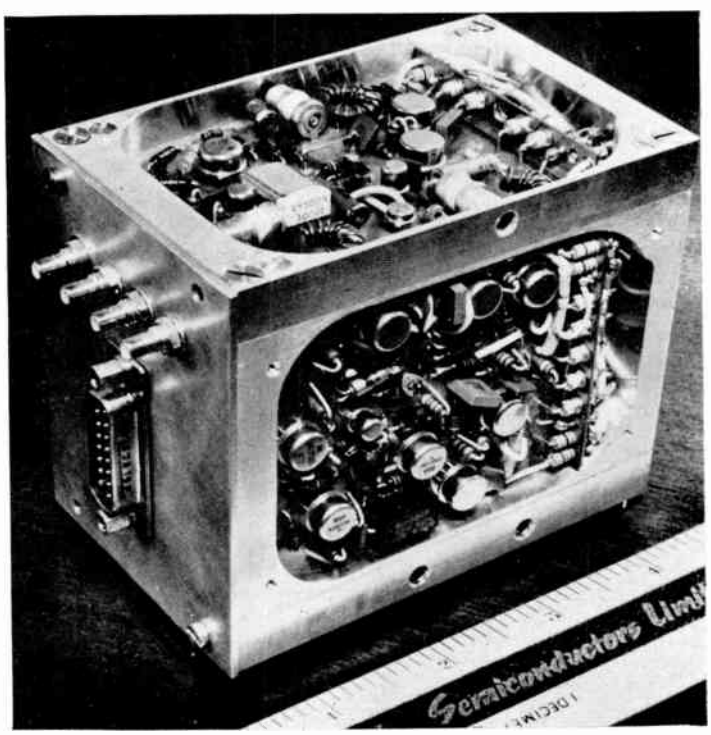


Fig. 8. Complete instrumentation with covers removed.

detector is removed by the filter consisting of R1,C1. IC1 was chosen for its low input bias current, the actual devices used being selected after temperature cycling tests for low temperature coefficient of input bias current, as this is an additional source of error. Setting of the zero crossover point is achieved by R6 and R9 whilst the gain of IC2 is given by

$$M_2 = \frac{R_{10}+R_7+R_8}{R_7+R_8}$$

As discussed earlier this gain is required to change by approximately 0·3% per degC in order to cancel out the term in the transfer function due to temperature. Resistor R_7 is therefore a positive temperature coefficient resistor which provides good thermal tracking over a wide temperature range. Figure 7 shows the error in output voltage versus input voltage for various temperatures, the error in most cases being less than the resolution of the telemetry system used, which is 1% of full scale.

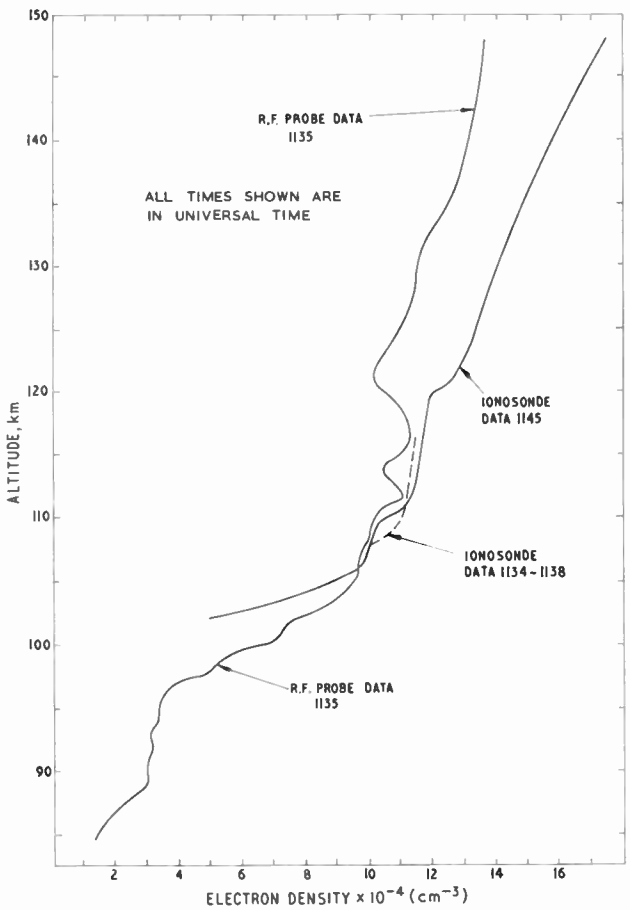


Fig. 9. Comparison of probe results with those of ionosonde.

5. Construction

Figure 8 is a photograph of the complete electronics package with covers removed. Four printed boards are used, mounted as shown. It will be seen that particular attention has been paid to filtering leads to each card to remove all unwanted coupling between circuits, especially at radio frequencies.

6. Performance of the Probe

The r.f. density probe has been used extensively in ionospheric investigations from rockets and satellites with very satisfactory results. In some cases it has been possible to correlate the results with those obtained from ground-based ionospheric soundings. Figure 9 shows such a comparison. The altitude scale shows true height as obtained from the rocket trajectory versus electron density as measured by the r.f. probe. Also shown are the ionosonde profiles, errors in height associated with ionogram reduction being approximately ± 3 km for heights up to 110 km increasing to ± 10–20 km at greater heights.

It will be seen that the agreement in absolute electron density is good within the limitations discussed above. The two peaks in the probe data are real and such peaks and troughs in the density give rise to ionosonde errors. This problem is discussed fully elsewhere.[7]

It may be concluded therefore that for *in situ* measurements the r.f. probe is capable of resolving detail in the ionospheric structure which is 'unseen' by an ionosonde based on the ground.

7. Acknowledgments

The authors would like to thank Professor J. Sayers for his active help and encouragement. The work was supported by a grant from the British Science Research Council.

8. References

1. Mackenzie, E. C. and Sayers, J., 'A radio frequency electron density probe for rocket investigation of the ionosphere', *Planet. Space Sci.*, **14**, pp. 731–40, August 1966.
2. Sayers, J., Wilson, J. W. G. and Loftus, B. T., 'Anomalous features of the electron density in the topside ionosphere', *Proc. Roy. Soc.*, **A311**, pp. 501–16, 1969.
3. Rothwell, P., Wager, J. H. and Sayers, J., 'Effect of the Johnston Island high-altitude nuclear explosion on the ionisation density in the topside ionosphere', *J. Geophys. Research*, **68**, No. 3, pp. 947–9, 1st February 1963.
4. Sayers, J., Rothwell, P. and Wager, J. H., 'Field aligned strata in the ionization above the ionospheric F2 layer', *Nature*, **198**, No. 4877, pp. 230–3, 20th April 1963.
5. Appleton, E. V., 'Wireless studies on the ionosphere', *J. Instn Elect. Engrs*, **71**, pp. 642–50, 1932.
6. Terman, F. E., "Radio Engineers' Handbook', Sect. 10, para. 8, p. 712 (McGraw-Hill, New York, 1943).
7. Goodall, C. V., 'Improved accuracy in in-situ probe measurements of ionospheric n_e and T_e using probe techniques requiring only low bandwidth telemetry channels', *Planet. Space Sci.*, **19**, pp. 827–40, August 1971.
8. Motorola Application Report, No. AN 261.
9. 'The Application of Linear Microcircuits', **1**, SGS-UK Ltd., 1969.

Manuscript first received by the Institution on 12th July 1971 and in revised form on 21st October 1971. (Paper No. 1448/AMMS48.)

U.D.C. 621.396.677

The Performance and Design of 2:1 Bandwidth Log-periodic Dipole Arrays

B. G. EVANS,
B.Sc., Ph.D., C.Eng., M.I.E.E.*

SUMMARY

Using the author's improved theory[7] the radiation pattern and impedance of a 2 : 1 bandwidth log-periodic dipole array are investigated over a wide range of design parameters. Using these results, a general design procedure is formulated for an array having a prescribed radiation characteristic and minimum input v.s.w.r. over the design band.

* *Department of Electrical Engineering Science, University of Essex, Colchester, Essex.*

1. Introduction

Log-periodic antennas have been the subject of considerable research and development since the original work of DuHammel and Isbell[1] showed them to possess frequency independent characteristics. Early devices were planar,[2, 3] but Isbell[4] showed that similar performances could be obtained with arrays of conventional elements and thus founded the log-periodic dipole array (l.p.d.a.). An important consequence of this advance was that the hitherto ill-understood log-periodic (l.p.) operation could be studied via the analysis of their operation in terms of conventional theory. Isbell laid the foundation for future designs but it was not until Carrel[5] produced his classical paper on the analysis of such structures that designs, albeit empirical, were possible. Carrel produced a mathematical analysis of the l.p.d.a. which took into account both the circuit and radiation aspects of the array. Although the radiation problem took into account the mutual coupling between dipole elements, it relied upon first-order sinusoidal theory which introduced inaccuracies into the analysis. Subsequently, more theoretical analyses[6, 7] have been performed on the structure which involve a complete array vector potential solution to the radiation problem, and which have resulted in more accurate predictions of experimental performance and a greater insight into the detailed operation of the array. It has been found that the general properties of the array as obtained from Carrel's results are not invalidated by the improved theories, over the design bandwidth, except at the band edges where the first order theory breaks down. Apart from this, there still remains the question of how the many array design parameters affect its performance. Thus whilst Carrel's results are still valid for an engineering approach to the problem, a more thorough investigation of the performance of the array, using the new theory, as its design parameters are varied, was considered a prelude to a more general design technique.

2. The Log-Periodic Dipole Array

As with all l.p. geometries, all dimensions of the dipole array are increased by a constant ratio in moving outwards from the origin. Thus lengths, spacings and diameters are related by a constant scale factor τ where, with reference to Fig. 1(a),

$$\frac{l_n}{l_{n-1}} = \frac{d_n}{d_{n-1}} = \frac{a_n}{a_{n-1}} = \tau$$

and the characteristic angle (note that l.p. antennas were derived from the equiangular variety) is represented by a space factor σ, where the convention used here is

$$\sigma = \tan\alpha = \frac{l_1}{d_1}$$

The active portion of the array (active region) from which most radiation occurs is centred around those elements near resonance (for which l_n is somewhat less than $\lambda/4$). As the frequency is changed, the active region moves back and forth along the array. Because practically all the input power is absorbed in and radiated by the

active region, the larger elements to the right of it are not excited and are in an essentially field-free region.

In order to produce practical l.p. antennas, the truly infinite structure must be replaced by a truncated one, affecting as little as possible its frequency independent behaviour. This means that at least one basic dimension needs to be specified, and this is usually the length of one of the extreme dipoles which will determine one of the array band edges. The array is usually fed, not at its true origin, but at the input to the first cell as shown in Fig. 1(a). Extending the geometry it will be seen that

$$D_0 = \frac{d}{(1-\tau)}$$

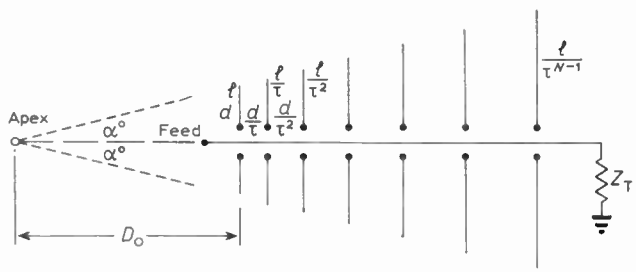
and

$$\sigma = \frac{l(1-\tau)}{d}$$

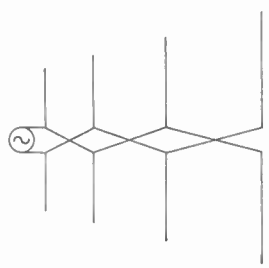
whence all distances d_n are expressible in terms of l, τ and σ. The bandwidth of the array is determined approximately by the resonant frequency of the extreme elements; hence for an N element array will be

$$\frac{1}{\tau^{N-1}} : 1$$

This will be degraded in practice by the active region running off the ends of the array before extreme element resonances. The difference between the two is called the active region bandwidth.



(a) Geometry,



(b) Feeding.

Fig. 1. Log-periodic dipole array.

The excitation of the structure has to be such as to produce endfire radiation in the direction of the apex, called backfire radiation. This is necessary because the farthest elements from the feed are resonant first, and as such any radiation in the endfire direction away from the feed would produce a field which would induce currents on the above resonance elements, which would in turn contribute to the far field and destroy the l.p. operation. For successful operation, the l.p.d.a. has been found to require feeding with a transposition of the transmission line between adjacent dipoles as shown in Fig. 1(b). This condition is sufficient but not necessary as several successful unswitched arrays have been produced[8] which derive the delay by different means. The antenna is thus caused to radiate in the backfire direction, a condition which is necessary for successful uni-directional frequency independent l.p. operation.

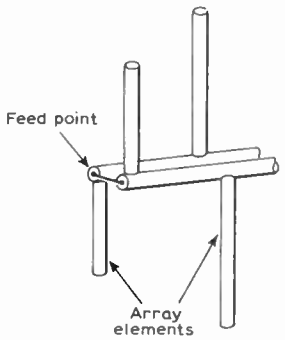


Fig. 2. Method of feeding dipoles with a transposed feeder.

As adjacent dipole currents near to the input of the array are nearly out of phase and due to the dipoles being close together, the radiated fields produced by them almost cancel each other. As the spacing d_n expands, a point is reached where the phase delay in the transmission line combined with the 180° switch gives a total of 360° $(1-d_n/\lambda)$. This puts the radiated fields from the dipoles in phase in the backward direction. Moving further out increases the phase delay and so the in-phase direction moves from backward through broadside to forward. Therefore a good beam off the apex should result if the dipoles are resonant when the total phase delay between dipoles is 360° $(1-d_n/\lambda)$.

The practical configuration of a l.p.d.a. in which transposition is used to create the slow wave condition in the feeder is shown in Fig. 2. This incorporates a frequency independent balun to convert from the coaxial feed to the biaxial transmission line on which the dipoles are positioned. In order to avoid interaction with radiated fields, the coaxial feeder is run inside one of the hollow cylindrical conductors, its outer and inner conductors being connected to the two cylindrical conductors at the feed point. Providing the separation between the biaxial conductors is less than $\lambda/16$, satisfactory transition is obtained between balanced and unbalanced conditions. The structure is terminated by a short-circuit plate whose dimensions are sufficient to short the majority of the biaxial line feed, and which is located beyond the longest element.

The basic design parameters are τ, σ, N and l, Z_f the feeder line characteristic impedance and the dipole l/a ratio. These affect the basic performance characteristics of radiation pattern and input impedance/v.s.w.r.—the question, so far unanswered, is in what way.

3. Performance Characteristics

A comparison between the results obtained using Carrel's theory and the improved theory show that there

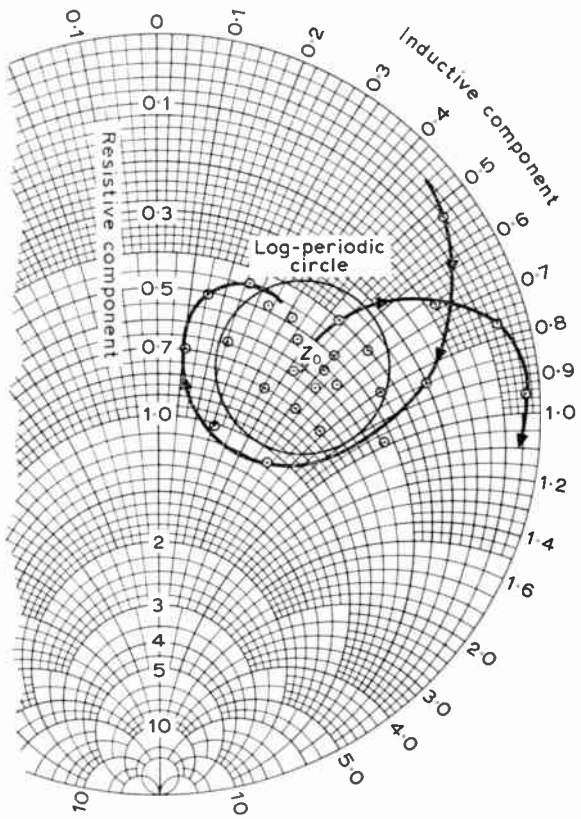


Fig. 3. Impedance plot of a 10-element l.p.d.a. ($\tau = 0{\cdot}9162$, $\sigma = 0{\cdot}6$).

is no significant differences between the general shape of the characteristics. However, at the band edges the results obtained from the new theory show much better agreement with experiment than the old, both for radiation pattern and v.s.w.r. responses.[7] The improvements in pattern are those of detail especially in side-lobe level and back-lobe predictions, whereas the impedance response shows much better agreement throughout the band with experiment and shows up the shortcomings of the old theory in predicting impedances.

The following are the results of an extensive experimental and theoretical investigation into the effects of the design parameters on the performance of a 2 : 1 bandwidth l.p.d.a.

3.1. *General Characteristics*

The input impedance of the array is shown plotted on a Smith chart in Fig. 3, from which the l.p. region is clearly defined. It will be noticed that at the low-frequency end the plot spirals around in a clockwise manner with frequency which is characteristic of an isolated dipole. This corresponds to the negative reactance of the under-resonant longest dipole. As the first dipole is approached, the plot spirals in to its approximate isolated value. If it were merely a discrete dipole, the plot would spiral out with positive reactance values. However, the active region of the array quickly encompasses the next dipole resonance and the plot spirals back again to the second dipole isolated impedance, which should in theory be the same as the first as the dipoles are electrically equal. This process continues as the active region moves along the array and is characterized by the l.p. circle shown in Fig. 3. When the active region runs off the small end of the array, the plot spirals outwards in a manner characteristic of the positive reactance of the above resonance smallest dipole. A measure of the smoothness of transition between the dipoles of the active region can be obtained from the size of the l.p. region, which should ideally be as small as possible for matching purposes.

The electromagnetic field of the array is determined by the currents in the elements. The minor-lobe structures in particular are very sensitive to the phase distribution along each element. The currents in the below resonance dipoles are relatively small but the new theory shows that their cumulative contribution to the far field pattern is comparable to that of parasitic directors in a Yagi array. They are certainly shown to contribute to the minor lobe structure and to the main beam width.

Far-field patterns of a 200–400 MHz array are shown in Fig. 4. The unilateral back-fire patterns are seen to be quite constant over the design range with slightly deteriorating shape, exemplified by the sensitive minor lobe, towards the high frequency end of the band. This is due to the different interactions with varying numbers of under resonant elements. The E-plane beamwidth is in general narrower than the H-plane and is characterized by a smaller minor lobe. At frequencies below 200 MHz the field patterns degenerate to those of an isolated dipole and at frequencies above 400 MHz the minor lobe becomes very large and the unidirectional property of the pattern is lost.

Initial investigations with other bandwidth arrays have shown this and other performances to be typical. Thus although only the 2 : 1 bandwidth results are presented herein, data which have already been obtained on wider bandwidth arrays would seem to confirm the more general applicability of the results.

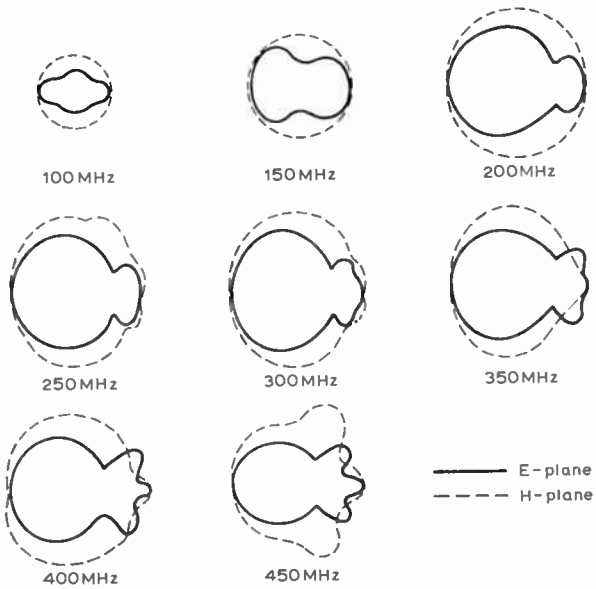


Fig. 4. Radiation patterns of a 10-element l.p.d.a., bandwidth 200–400 MHz.

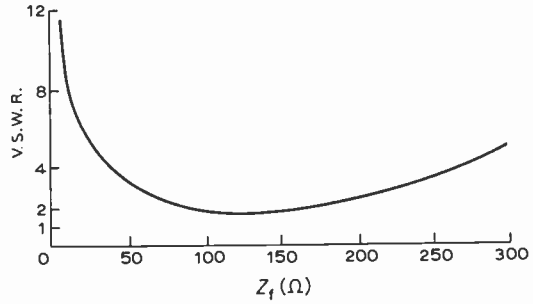


Fig. 5. Variation of optimum input v.s.w.r. with feeder impedance.

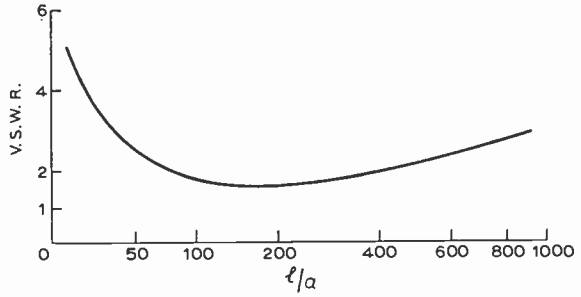


Fig. 6. Variation of optimum input v.s.w.r. with dipole l/a ratio.

3.2. *Active-region Bandwidth*

It will be noticed from the characteristics presented so far that the design bandwidth is degraded at both ends (205–385 MHz). This is due to the active region running into the truncations. When this happens the active region is no longer constant in terms of wavelengths, which is one of the basic conditions for l.p. operation, and so the l.p. operation of the array breaks up before the extreme dipole resonances are reached. The asymmetrical nature of the break-up is again due to the differing contributions to the active region.

Investigations have shown that the active-region bandwidth is dependent upon many parameters including τ, σ, N and l/a, in short any parameter that affects either the element spacing or the transition of the active region between elements. Unfortunately, the dependence upon these parameters is very complex and it is not possible to develop an analytic expression for the active-region bandwidth. However, it was found from a study of the extensive performance characteristics, that by specifying a slightly higher upper-band frequency and decreasing τ correspondingly, the active region bandwidth could be compensated by using the empirical formulae:

$$f_{\text{upper}} \rightarrow 1{\cdot}1 \times f_{\text{upper design}}$$

Therefore,

$$\tau \rightarrow \tau_{\text{design}} \times 10 \frac{(0{\cdot}0414)}{N-1}$$

3.3. *Transposed Feeding*

To physically-realize the 180° phase delay between elements by a switched feeding method it is necessary to have a transverse displacement between the two halves of the dipoles as shown in Fig. 2. The characteristics of such a transversely displaced dipole differ from those of the coplanar structure, but this was found to affect array performance only if the feeder separation D was greater than $0{\cdot}02\lambda_{\text{min}}$.[9] Another effect of utilizing separations greater than $0{\cdot}02\lambda_{\text{min}}$ and not increasing the feeder displacement periodically by τ as true l.p. geometry dictates is a rotation in the radiation pattern with frequency. All such effects may be eliminated by using a constant feeder-line displacement of less than $0{\cdot}02\lambda_{\text{min}}$.

† The input referred to is the true array feed point and does not include the length of coaxial feeder line.

3.4. *Input Impedance/V.S.W.R. Dependence on Design Parameters*

The dependence of array bandwidth on the factor τ^{N-1} has already been shown. The independent determination of τ and N is very much more arbitrary as far as impedance properties are concerned. However, general rules which do apply are that τ should be kept as close to unity as possible in order to minimize variations in the band and N should be large enough to provide adequate overlapping or contiguous active regions. It is interesting to note that the active-region correction is approximately $1/N$ of the design bandwidth which could indicate that the active regions are more or less contiguous, which would be a feature of desirable operation. Too many elements and the high frequency performance shows an increased oscillatory performance.

The characteristic impedance of the feeder, Z_f, largely determines the level of the input impedance and v.s.w.r. of the array. The variation of optimum input v.s.w.r. with Z_f is shown in Fig. 5, from which it can be seen that there is a region between 75 and 150 Ω in which the v.s.w.r. is at a minimum. It also shows that input v.s.w.r.s of the order of 2 : 1 are common with this structure.† The individual dipole impedance does govern the v.s.w.r. and therefore the l/a ratio would be expected to have some control over it. Variation of the optimum input v.s.w.r. with l/a ratio is shown in Fig. 6 to exhibit a minimum value between 100–200. The centre of the l.p. circle on the Smith chart also rotates as the l/a ratio or Z_f are increased. The positioning of the short circuit termination is relatively unimportant to the impedance characteristics. This is due to the rapid decay of currents on the array elements past the active region. However, at the low-frequency band edge the placing can affect the characteristics and it was found that a distance of $\lambda_{\text{max}}/8$ from the longest dipole was sufficient for these effects to be negligible.

Although the bandwidth of the array is fixed by the τ^{N-1} ratio it was found that the spacing factor σ did affect it slightly. In general an increase in σ produces an increase in the v.s.w.r. level by an amount which depends upon the number of dipoles. An increase in σ means an increase in the array angle, so that the elements are bunched together and have increased mutual interactions. Obviously this effect is more pronounced for larger N and so the choice of σ and N cannot be divorced; in general the larger N, the smaller the choice

of σ and vice versa. It has been found in practice that it is not advisable to use angles outside the range 20–60° and that the most favourable range is 25–45° ($\sigma = 0{\cdot}4$–$1{\cdot}0$). For example, the 2 : 1 bandwidth array with $N = 10$ has $\sigma = 0{\cdot}6$ for best operation, $N = 20$, $\sigma = 0{\cdot}45$ and $N = 5$, $\sigma = 0{\cdot}75$. The variation of the l.p. centre, i.e. the mean input impedance, with σ and N is shown in Fig. 7. The resistive as well as reactive components decrease with increasing σ indicating reduced array contributions. Also both components decrease with the addition of more elements although the reactive portion decreases faster. With a large number of elements, the overlapping of the active regions is accentuated and the reactances associated with transitions are reduced. This leads to a predominantly resistive input which is desirable for matching purposes. The variation of input impedance with σ is quite small, the only noticeable effect being a change in the mean levels as shown in Fig. 7.

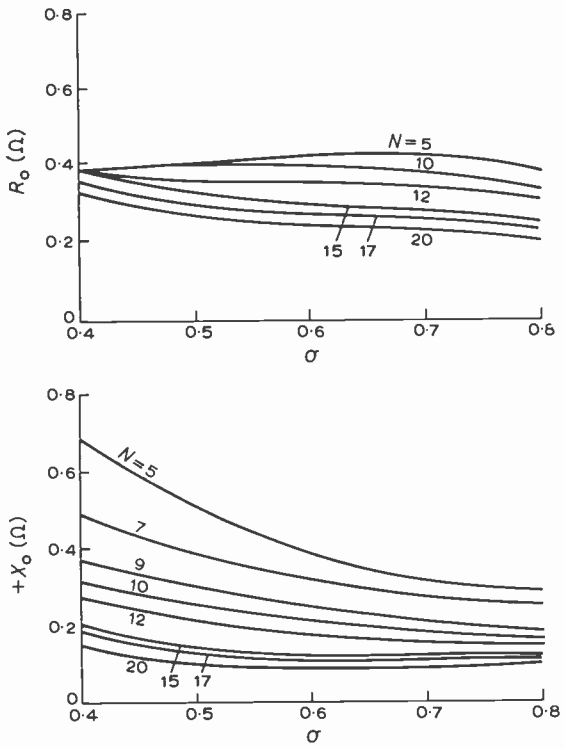


Fig. 7. Variation of mean input impedance with spacing factor σ. R_0, X_0 is the centre of l.p. circle on the Smith chart. $Z_f = 100\ \Omega$; $l/a = 120$.

For larger bandwidths it was seen that with too few array elements, the l.p. region degenerated to an ellipse. It was further noticed that the lower reactive values occurred at frequencies below mid-band and the higher above mid-band, whilst the resistive values remained essentially constant. This elliptical nature of the impedance plot is characteristic of too few array elements or too low a choice of σ. When this occurs, the transitions of the active region between resonances become discontinuous, giving rise to larger off-resonance reactive components which distort the circle in the reactance plane. So that if characteristics of this nature are encountered N or σ, or both need to be increased to map the plot into the circle which characterizes good l.p. operation.

It was noticed throughout the investigation that the average array input impedance Z_0 was less than the feeder impedance Z_f. This, together with the fact that the active region gave small v.s.w.r. values over the band, suggests that the array input impedance may be approximated by the feeder line capacitively loaded by the small dipole elements as in Carrel,[5] but that this is not an accurate method of determining the average input impedance due to the interactions.

3.5. *Radiation Pattern Dependence on Design Parameters*

The radiation pattern has already been seen to remain tolerably constant over the design bandwidth and is subject to the same active region degeneration as the impedance characteristics. What remains to be determined is the variation of the pattern characteristics, gain, side-lobe level, front-to-back ratio and beamwidth with the design parameters.

The separate variation of the pattern characteristics with differing values of τ and N, but still subject to a constant τ^{N-1} bandwidth are shown in Fig. 8. All characteristics behave in a similar way in oscillating at low values of N and τ but gradually smoothing out and levelling off to a pseudo-saturation level, above which the addition of extra elements provides little variation in the response. For the case of the 2 : 1 bandwidth array it is obvious that the inclusion of more than 13 elements adds insignificantly to the performance. It will also be noted that maximum beamwidth and gain occur concurrently with low side-lobes and maximum front-to-back ratio. For the array with $\sigma = 0{\cdot}6$ and having 13 elements, the E-plane beamwidth is 68·5°, the H-plane beamwidth 133°, gain 6·6 dB and front-to-back ratio 17 dB. There will obviously be correspondingly shaped responses for other bandwidths with different threshold values of N.

Parameters N and τ may thus be chosen to optimize the pattern performance, but if the resulting performance still does not meet design requirements, σ remains to make further adjustments. The spacing factor σ controls the added array phase differences between elements which determines the element phase distribution, and the variation of pattern characteristics with it are shown in Fig. 9. Whence σ has the desirable quality of adjusting the basic levels of responses whilst leaving their shape unaltered. From Fig. 9 it will be seen that larger beamwidths are accompanied by increased back-lobes and decreased gain, hence whilst the choice of σ must be balanced against the undesirable side-lobes. Using both σ and τ it is possible to design for a range of differing responses.

4. The Design of L.P.D.A.s

It is clear from the information obtained in the last section that the parameters τ, σ and N affect the pattern characteristics to a much greater extent than the non-pattern affecting parameters Z_f and l/a. Bearing in mind

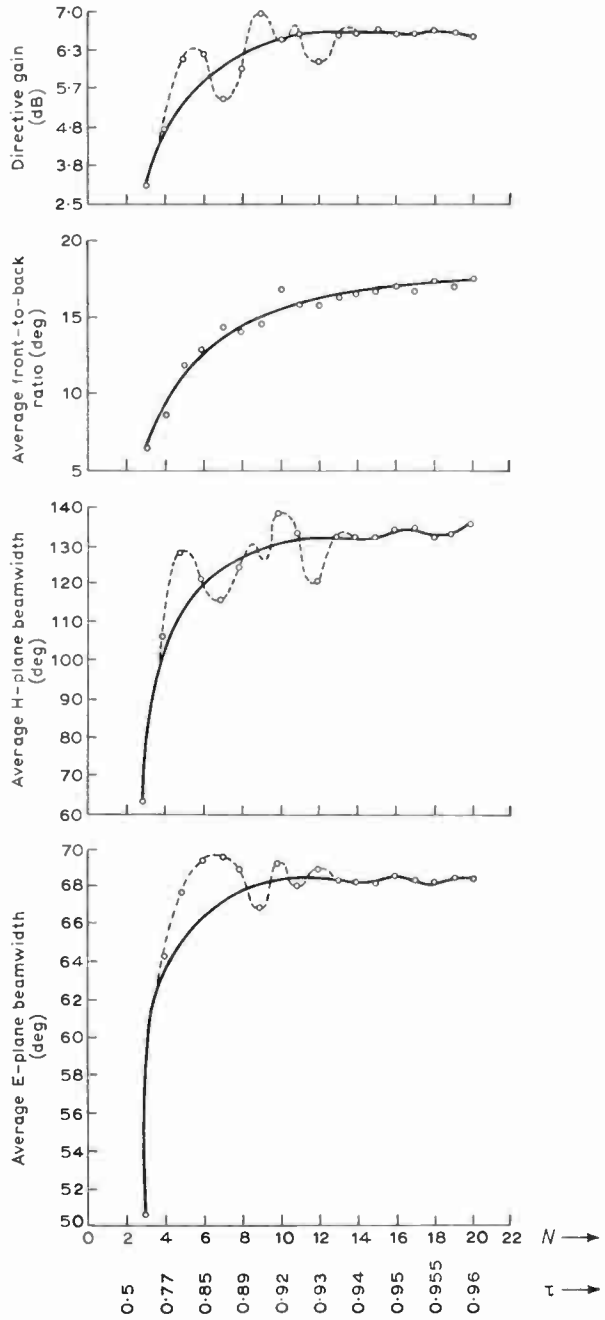


Fig. 8. Variation of pattern characteristics with N and τ. ($Z_f = 100\ \Omega$; $l/a = 120$.)

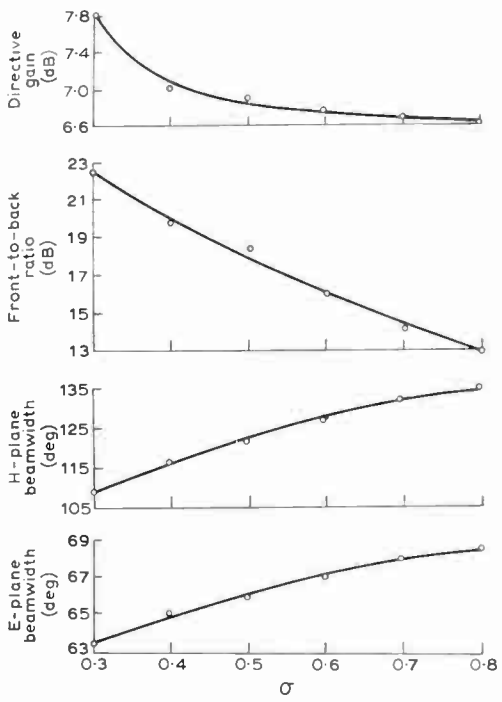


Fig. 9. Variation of pattern characteristics with σ. ($Z_f = 100\ \Omega$; $l/a = 120$.)

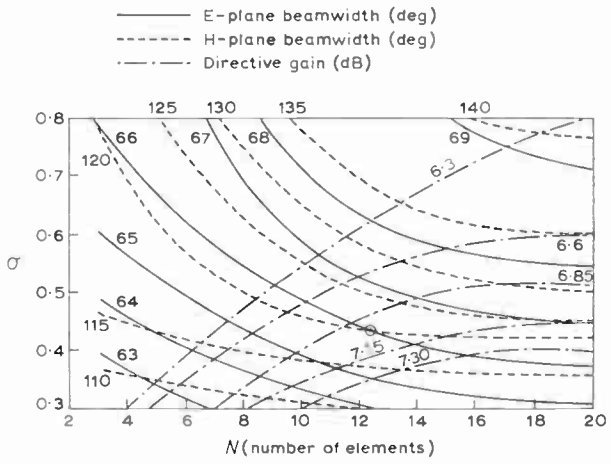


Fig. 10. Radiation pattern synthesis chart for 2 : 1 bandwidth l.p.d.a.

that from a design point of view the patterns are of greater importance than the impedance level which could possibly be compensated, it will be seen that the pattern and impedance designs may be conveniently separated. It is worth noting that this is an approximation as there are small interactions between the two, especially via σ which affects the v.s.w.r. level slightly. The philosophy of design then is to choose the parameters N, τ and σ for required pattern characteristics and with these to design for minimum v.s.w.r. by appropriate choice of Z_f and l/a.

The design bandwidth immediately fixes τ^{N-1} and this may be split up together with a choice of σ, according to the pattern requirements which have been redrawn in Fig. 10. This process will require some juggling with parameters until acceptable responses are achieved. It is worth remembering that σ governs the boom-length of the array and thus in order to achieve a compact structure, some consideration must be given to selecting a high value of σ. Once the choice for the values of τ, σ and N is made, the active-region correction may be applied to give the final pattern forming parameters.

The design information relevant to the minimization of v.s.w.r. is presented in Fig. 11, with the exception of l/a bounds already discussed. This shows the impedance of the biaxial feeder Z_f as a function of radius and

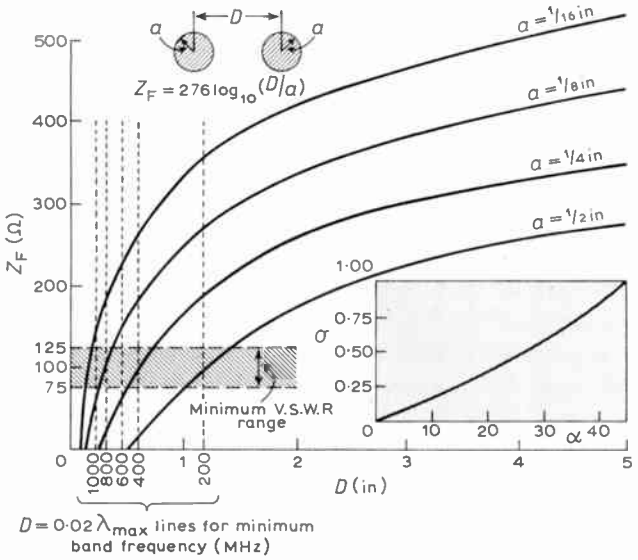


Fig. 11. Impedance/v.s.w.r. synthesis chart for 2 : 1 bandwidth l.p.d.a.

separation of the conductors. Superimposed on this is the range of Z_f which has been found to give minimum v.s.w.r.s (shaded area) and the limiting ordinates of $D = 0{\cdot}02\lambda$ (dotted lines) for negligible transverse feeder effects. A trial D and a are selected from the design box, bounded by the limiting impedance lines and the relevant D ordinate. The maximum dipole length l_N is then calculated and a l/a ratio chosen so that it lies within the recommended range and so that the maximum dipole diameter does not exceed the feeder conductor diameter. If such a choice is not possible, another value of D and a must be chosen from the design box and the process repeated until acceptable values are obtained.

Using the above design procedure together with the construction details outlined earlier a l.p.d.a. may be designed and constructed to have prescribed radiation characteristics and having a minimum input v.s.w.r. Although only the 2 : 1 bandwidth design curves are presented here, the computer program is available for the computation of curves for any desired bandwidth, the design procedure being the same in each case.

An example of the design procedure follows:

Specification

E-plane beamwidth 65°

H-plane beamwidth 120°

Gain not less than 6·5 dB

Front-to-back ratio better than 15 dB down on main lobe

V.s.w.r. tolerable over a 2 : 1 bandwidth centred on 300 MHz is 2 : 1

Design

Nominal bandwidth 200–400 MHz

$$\tau^{N-1} = 0{\cdot}5$$

(a) Pattern characteristics:

From Fig. 11 E-plane beamwidth 66°
H-plane beamwidth 120°
Gain 6·9 dB
} $N = 12$, $\sigma = 0{\cdot}44$

From Fig. 10 $N = 12$, $\sigma = 0{\cdot}44$ } front-to-back ratio 18·5 dB

Boom length $= 1{\cdot}295\lambda_{min}$ ($\sigma = 0{\cdot}44$)
$= 38{\cdot}2$ in $\simeq 1{\cdot}0$ m

$\alpha = 24°$.

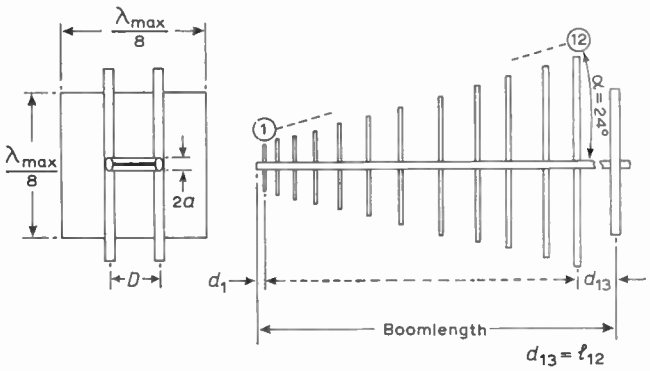
For $N = 12$, $\tau = 0{\cdot}9389$.

Active-region correction $\tau = 0{\cdot}9389 \times 10^{\frac{(0{\cdot}414)}{11}}$

$\tau = 0{\cdot}9308$.

(b) Impedance characteristics:

With a minimum band frequency of 200 MHz choose $D = 0{\cdot}265$ in (6·4 mm), $a = 0{\cdot}125$ in (3·175 mm) from the box. This gives $Z_t = 110\ \Omega$; choose $l/a = 120$ so that by reference to Figs. 5 and 6, v.s.w.r. is 2·0.



Element number	Spacing d (in)	Half lengths (in)	Diameters (in)
1	1·180	7·375	0·062
2	1·268	7·924	0·066
3	1·362	8·514	0·071
4	1·464	9·146	0·076
5	1·572	9·828	0·082
6	1·688	10·558	0·088
7	1·814	11·344	0·095
8	1·948	12·188	0·102
9	2·094	13·095	0·109
10	2·250	14·070	0·117
11	2·418	15·116	0·126
12	2·592	16·240	0·135

N	τ	σ	Z_t	l/a	D (in)	a (in)
12	0·9308	0·44	110	120	0·265	25

Fig. 12. Design specifications for 2 : 1 bandwidth l.p.d.a.

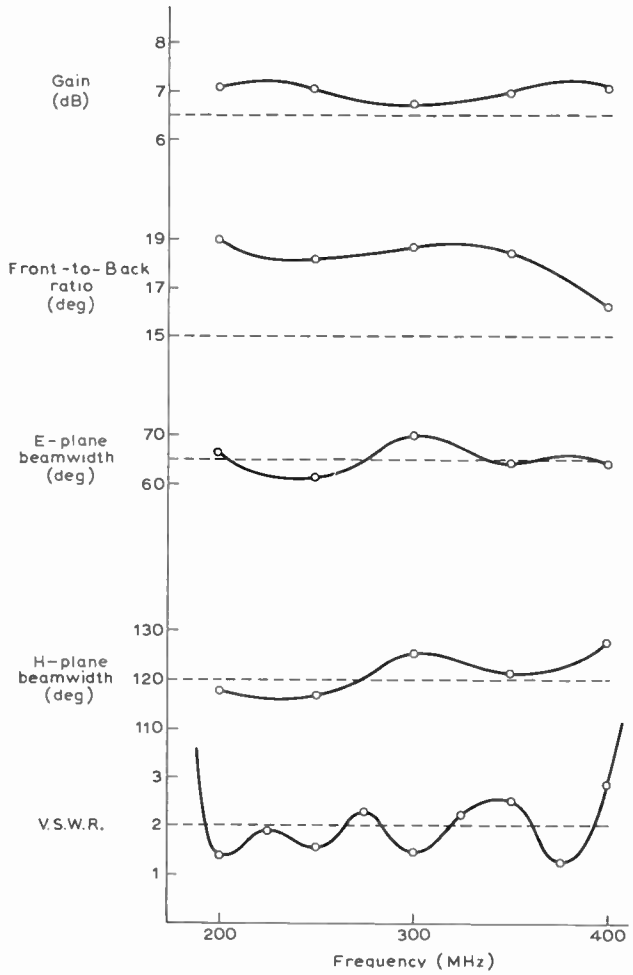


Fig. 13. Experimental performance of 12-element 2 : 1 bandwidth l.p.d.a.

To check that these are within the tolerance of working limits, extreme dipole diameters are ($d_n = 2a_n$),

$$l_1 = 7{\cdot}373 \text{ in } (18{\cdot}725 \text{ cm})$$

$$d_1 = \frac{7{\cdot}373}{60} = 0{\cdot}062 \text{ in } (1{\cdot}58 \text{ mm})$$

$$l_{12} = 16{\cdot}24 \text{ in } (41{\cdot}275 \text{ cm})$$

$$d_{12} = 0{\cdot}135 \text{ in } (0{\cdot}364 \text{ mm})$$

Both are physically realizable and $d_{12} < 2a$ so the design is acceptable. A full design chart is shown in Fig. 12. This design was built and tested to show the validity of the procedure. The experimental results are shown in Fig. 13.

5. Conclusions

Extensive data have been provided on the performance variation with design parameters of a 2 : 1 bandwidth l.p.d.a. As a result of this a design philosophy has been proposed which allows radiation as well as impedance characteristics to be included. Although only the 2 : 1 bandwidth array results are presented this is a fairly important and extensively used case and as such the results should prove useful to designers in the field. Design information on wider bandwidth arrays will be the subject of future papers.

6. References

1. DuHammel, R. H. and Isbell, D. E., 'Broadband logarithmically periodic antenna structures', *Inst. Radio Engrs Nat. Conv. Rec.*, **5**, Pt. 1, pp. 119–28, 1957.
2. Rumsey, V. H., 'Frequency independent antennas', *I.R.E. Nat. Conv. Rec.* **5**, Pt. 1, pp. 114–8, 1957.
3. DuHammel, R. H. and Ore, F. R., 'Logarithmically periodic antenna design', *I.R.E. Nat. Conv. Rec.*, **6**, Pt. 1, pp. 139–51, 1958.
4. Isbell, D. E., 'Log-periodic dipole arrays', *Trans. I.R.E. on Antennas and Propagation*, **AP-18**, pp. 260–7, May 1960.
5. Carrel, R. L., 'The design of log-periodic dipole antennas', *I.R.E. Int. Conv. Rec.*, **9**, Pt. 2, pp. 61–75, 1961.
6. Cheong, W. and King, R. W. P., 'Log-periodic dipole antennas', *Radio Science*, **2**, pp. 1315–25, November 1967.
7. Evans, B. G., 'Potential integral theory for a log-periodic array of N, parallel, non-staggered elements', *The Radio and Electronic Engineer*, **39**, pp. 224–82, April 1970.
8. Berry, D. G. and Ore, F. R., 'Log-periodic monopole arrays', *I.R.E. Int. Conv. Rec.*, **9**, Pt. 1, pp. 76–85, 1961.
9. Evans, B. G., 'The effects of transverse feed displacements on log-periodic dipole arrays', *Trans. I.E.E.E.*, **AP-18**, pp. 124–8, January 1970; also **AP-18**, p. 728, September 1970.

Manuscript first received by the Institution on 11th August 1971 and in revised form on 7th October 1971. (Paper No. 1449/Com 53.)

U.D.C. 621.396.67

Height Reduction Factor of Amplifier Loaded Antennas

P. A. RAMSDALE, B.Sc.*

and

T. S. M. MACLEAN,
Ph.D., C.Eng., M.I.E.E.†

SUMMARY

The problem of matching a short monopole antenna by incorporating an ideal amplifier in the monopole structure is analysed. It is shown that this leads to a simple method of predicting which types of transistor amplifier configuration are likely to lead to a height reduction factor over the $\lambda/4$ passive case.

* *Electrical Engineering Department, Lanchester Polytechnic, Coventry, formerly at the University of Birmingham.*

† *Department of Electronic and Electrical Engineering, University of Birmingham, Birmingham B15 2TT.*

1. Introduction

In a recent paper[1] expressions were derived for the properties of loop-monopole antennas containing transistors. It was shown that for some orientations of the transistors the antenna became resonant when considerably shorter than the $\lambda/4$ length of the conventional passive monopole, and unlike the tuned monopole the bandwidth was not reduced. However, the expression that was derived for the height reduction factor is extremely complicated, and hence it is difficult to isolate the features responsible for producing it.

In this paper the resonant height of a monopole containing an idealized amplifier will be found. From this result it will be shown that it is possible to predict in some measure the types of active loop dipole antennas which possess a resonant height reduction property.

2. Input Reactance of Monopole Antenna incorporating Ideal Series-connected Amplifier

Using the trigonometric current distribution approximations, the currents and potentials on the antenna in Fig. 1(a) may be written as

$$I_1(z) = C\cos kz + D\sin kz \qquad \text{.....(1)}$$

$$I_2(z) = A\sin k(2h-z) \qquad \text{.....(2)}$$

$$V_1(z) = -\mathrm{j}Z_c(C\sin kz - D\cos kz) \qquad \text{.....(3)}$$

$$V_2(z) = -\mathrm{j}Z_c A\cos k(2h-z) \qquad \text{.....(4)}$$

where A, C and D are constants, and Z_c is the average characteristic impedance of the monopole.

Fig. 1. (a) Active transmitting monopole antenna with amplifier at mid-point.
(b) Active transmitting monopole as in (a) with amplifier reversed.

The boundary conditions imposed by the amplifier are

$$I_2(h) = G_I I_1(h) \qquad \text{.....(5)}$$

and

$$V_2(h) = G_V V_1(h) \qquad \text{.....(6)}$$

where G_I, G_V are the current and voltage gains of the ideal amplifier.

The input impedance of the antenna is given by

$$Z_i = \frac{V_1(0)}{I_1(0)} = \mathrm{j}Z_c\frac{D}{C} \qquad \text{.....(7)}$$

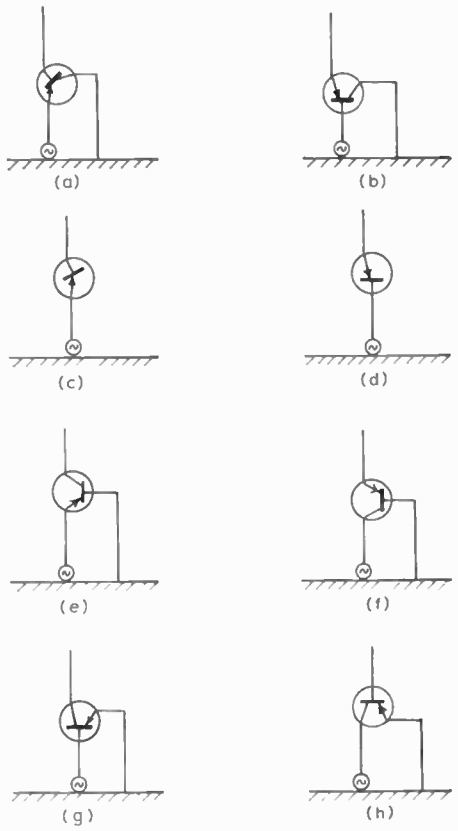
and applying the boundary conditions at the amplifier gives for the input reactance

$$X_i = -\mathrm{j}Z_c\frac{G_I\cos^2 kh - G_V\sin^2 kh}{(G_I+G_V)\sin kh\cos kh} \qquad \text{.....(8)}$$

The sign of the input reactance changes when

$$\cos kh = 0 \qquad \text{.....(9)}$$

$$\sin kh = 0 \qquad \text{.....(10)}$$



(a) Active loop-monopole antenna in fed-emitter, collector-loop (f.e.c.l.) configuration.
(b) Active loop-monopole antenna in fed-base, collector-loop (f.b.c.l.) configuration.
(c) Active monopole approximation to f.e.c.l.
(d) Active monopole approximation to f.b.c.l.
(e) Active loop-monopole antenna in fed-emitter, base-loop (f.e.b.l.) configuration.
(f) Active loop-monopole antenna in fed-collector, base-loop (f.c.b.l.) configuration.
(g) Active loop-monopole antenna in fed-base, emitter loop (f.b.e.l.) configuration.
(h) Active loop-monopole antenna in fed-collector, emitter loop (f.c.e.l.) configuration.

Fig. 2.

and

$$\tan kh = \sqrt{\frac{G_I}{G_V}} \qquad \text{.....(11)}$$

The first two solutions give resonances or antiresonances at monopole heights

$$2h = \frac{\lambda}{2}, \ \lambda, \ \frac{3\lambda}{2} \ldots \qquad \text{.....(12)}$$

while it can be seen from the third solution that if $G_I < G_V$ then the resonant monopole antenna is shorter than $\lambda/4$.

The analysis may be repeated for the case in which the amplifier is reversed as in Fig. 1(b). The solution of eqn. (11) then becomes inverted so that for height reduction $G_V < G_I$.

3. Application to 3-terminal Transistor Amplifiers incorporated in Loop Monopole Antennas

It has been shown previously[1] that the input impedance of an active antenna is the same when transmitting as when receiving. This follows from the principle of superposition. Consequently when calculating input impedance it is sufficient to consider the transmitting case only. This is simpler since

(i) to a first approximation, there is a signal applied to only one section of the antenna

and (ii) the amplifier gains are not, as in the receiving case, dependent on the antenna terminal load impedance.

It must be realized that the calculation of the exact antenna impedance is not straightforward as there are small but finite self and mutual impedances which can be considered as additional impedance loads on the amplifier incorporated in the aerial. However, it is not the purpose of this paper to carry out involved calculations of this kind, but rather to present a simplified solution of a very complex problem.

In order to make comparisons between loop monopole antennas and the amplifier-loaded monopoles, further approximations are necessary:

(1) The effect of the loop wire is ignored so that the loop monopole reduces to the same form as the loaded monopole of Section 2. It is to be expected that this will result in least inaccuracy when the loop is connected to the collector of a transistor amplifier, because of the large incremental impedance associated with this electrode, but this simplification is also applied to the other configurations in Sections 3.2 and 3.3.

(2) The transistor equivalent circuit representation will include several passive lumped components and these will also influence the resonant properties of the antenna.[2]

3.1. *The Fed-emitter Collector Loop (f.e.c.l.) and Fed-base Collector Loop (f.b.c.l.) Active Antennas*

The network configurations for these are shown in Figs. 2(a) and (b). Approximation (1) means that these circuits will be treated as shown in Figs. 2(c) and (d), where both amplifiers are to be considered in terms of

common collector operation. In Fig. 2(c) the amplifier has an emitter input so that current gain $\simeq 1/\beta$, whereas in Fig. 2(d) with a base input, current gain $\simeq \beta$. On the other hand for the common collector, i.e. emitter follower, configuration, typically voltage gain ~ 1. Hence from eqn. (11),

$$\tan kh \simeq \frac{1}{\sqrt{\beta}} \qquad \text{(f.e.c.l.)}$$

and

$$\tan kh \simeq \beta \qquad \text{(f.b.c.l.)}$$

giving a height reduction factor for the f.e.c.l. case, since $\beta > 1$ and therefore not for the f.b.c.l. case.

Typical numerical values of the height reduction factor are shown in Fig. 3 as a function of the transistor β and are compared with those for a typical low-frequency transistor. These have been obtained using the input impedance expressions from reference 1 and it can be seen that the two results are very similar. The third curve in this diagram is Meinke's solution[3] for the same antenna, which was found by using a lumped circuit representation for the radiating elements.

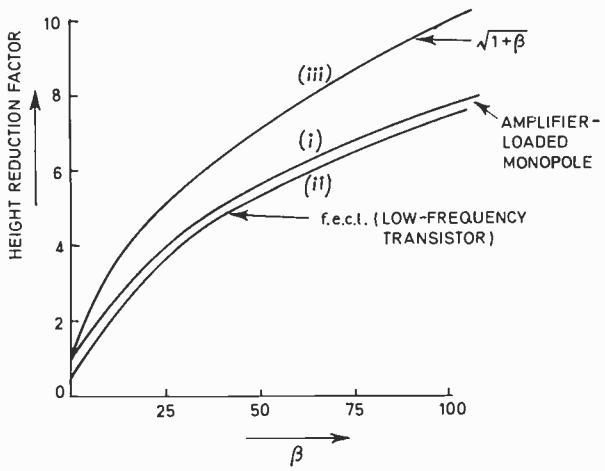


Fig. 3. Height reduction factor of f.e.c.l. antenna as function of transistor β.

(i) Results from amplifier-loaded monopole approach.

(ii) Results from reference 1.

(iii) Results from reference 3.

3.2. *The Fed-emitter Base Loop (f.e.b.l.) and Fed-collector Base Loop (f.c.b.l.) Active Antennas*

For the f.e.b.l. and f.c.b.l. networks shown in Figs. 2(e) and 2(f) the analysis is considered in terms of the common base configuration with emitter and collector inputs respectively. Approximation (1) means that in this case the lead in the base circuit is ignored as far as the input reactance is concerned. The current gains are both of the order of unity. The voltage gain is large for the emitter input f.e.b.l., and substantially less than unity for the collector input f.c.b.l., as would be expected for any collector input circuit. Hence there is a large height reduction factor for the f.e.b.l. and none for the f.c.b.l. Using the loop monopole analysis[1] with the equivalent-T transistor representation gives a height factor of typically 50 for the f.e.b.l.

3.3. *The Fed-base Emitter Loop (f.b.e.l.) and Fed-collector Emitter Loop (f.c.e.l.) Active Antennas*

The networks showing the f.b.e.l. and f.c.e.l. cases are given in Fig. 2(g) and 2(h). Here the current gains between base and collector on the one hand and between collector and base on the other are taken as β and $1/\beta$ respectively. The voltage gain of the common emitter f.b.e.l. is substantially the same as for the common base f.e.b.l. so that for resonance

$$\tan kh_{\text{f.b.e.l.}} \simeq \sqrt{\beta} \tan kh_{\text{f.e.b.l.}}$$

i.e. a resonant f.b.e.l. antenna is several times longer than the corresponding f.e.b.l. antenna. The loop monopole analysis of reference 4 does not in fact give a height reduction factor for a characteristic impedance of 200 Ω although more recent work indicates that raising Z_c does give a height reduction. On the other hand, variations in Z_c for the f.e.c.l. case have very little effect on height reduction, and for the f.e.b.l. case likewise.

The voltage gain of the f.c.e.l. amplifier with the input connected to the collector is extremely small as for the f.c.b.l. case in Section 3.2, but since the current gain is lower by a factor of $\sim \beta$ there is a greater possibility of a height reduction factor for comparable voltage gains. For the low-frequency T transistor circuit there is no resonant height reduction but reference 1 illustrates an example of a practical high frequency f.c.e.l. which exhibits this reduction property.

4. Conclusions

Using the ideal amplifier approach it appears that four of the six types of single transistor loaded loop monopoles may exhibit a height reduction property. It may now be possible to make useful predictions of the effects of multi-transistor loadings on active loop monopole antennas.

5. References

1. Ramsdale, P. A. and Maclean, T. S. M., 'Active loop-dipole aerials', *Proc. Instn Elect. Engrs*, **118**, No. 12, pp. 1698–1710, 1971.

2. Ramsdale, P. A. and Maclean, T. S. M., 'Control of resonant frequency of monopole aerial by resistive loading', *Intl. J. Electronics*, **31**, pp. 77–81, 1971.

3. Meinke, H., 'Aktive Antennen', *Nachrichtentech. Z. (N.T.Z.)* **19**, pp. 697–705, 1966.

4. Ramsdale, P. A., 'Active Loop-dipole Antennas'. M.Sc. Thesis, University of Birmingham, 1970.

Manuscript first received by the Institution on 21st December 1971 and in final form on 27th January 1972. (Short Contribution No. 154/Com 54.)

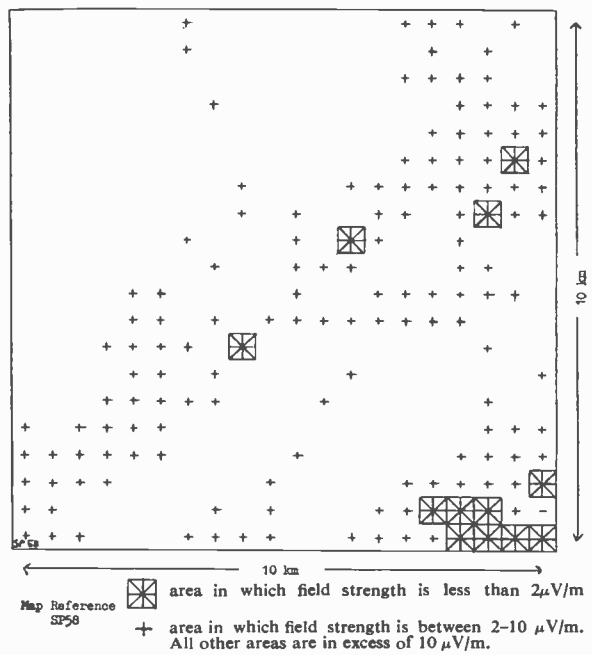
Computer-derived Mobile Radio Network Surveys

The electricity, gas and coal industries together comprise the largest user of private mobile radio communications in the United Kingdom, operating some 700 v.h.f. base stations and 25 000 mobile stations, together with about 350 point-to-point u.h.f. links, and substantial networks of microwave links. The co-ordination of the industries' requirements and the administration of these radio channels in co-operation with the Ministry of Posts and Telecommunications, is the responsibility of the Joint Radio Committee of the Nationalized Power Industries (JRC).

With the growing use of mobile radio systems in all sectors of industry and commerce, problems associated with frequency assignment and interference are a major consideration, since it is necessary to assign frequencies on as close a radio spectrum and geographical spacing as possible to optimize the use of available frequencies. Over recent years, public attention has been focused on the potential adverse effect of a proliferation of radio stations on the countryside, with the result that it has now become more difficult to establish a new radio station in the best location for the purpose. It is therefore more necessary than ever before to determine the radio coverage area from each of a number of possible radio transmitter sites and to select the best.

In carrying out a radio site survey, the normal practice is to calculate aerial height, the power to be transmitted and any aerial directivity properties, in relation to the terrain over which coverage is required. Certain limitations as to aerial height may be forced on the user by amenity considerations, and licensing regulations may restrict the transmitter power and aerial design. A temporary transmitter is then erected on the proposed site, and a field survey is conducted over the area, during which field strength evaluations are made at selected or random points. The survey is essentially concentrated on the areas in which marginal signal levels can be expected, and is generally confined solely to the required operational area. Such a survey may take many months to complete, although in the case of a private mobile radio system for industrial or commercial operations the survey would not usually extend beyond two weeks or so. Practical surveys of this nature are costly in equipment, manpower and time, and provide only a very limited amount of information. They also suffer from many practical disadvantages.

As a result of work initiated by Central Electricity Research Laboratories with the University of Manchester Institute of Science and Technology, the Joint Radio Committee of the Nationalized Power Industries have developed and put into operation a computer program which predicts the field strength and service area for a given radio site. The CEGB IBM360 computer is provided with data relating to the transmitter map reference, the height above ground and effective radiated power of the transmitter aerial, the type of mobile receiver aerial and its height above ground, and the area to be surveyed. The computer output comprises a print-out of calculated field strengths and signal levels at half-kilometre intervals for the area surveyed, and in addition if needed, a transparency plot suitable for overlay on Ordnance Survey 1 in/1 mile (1:63360), 2½ in/1 mile (1:25000) or ¼ in/1 mile (1:253440) maps as required. The plot provides a clear and easily understood picture, when overlayed on the appropriate OS map, of the comparative levels of signal over the area. The accompanying figure shows a 10 km square section of a survey originally plotted at a scale of 1 in/1 mile.



Sample plot for 10 km square section of a survey

Since a knowledge of the topography of the area under survey is called for in the calculation of radio field strengths at any given point, it has been necessary to create a data base of ground height. The work of extracting this information was carried out by PMA Consultants Ltd. of Horley, who assembled a data bank of 880 000 points at ½ km intervals for the whole of England, Wales and Scotland excepting the Highlands and Islands. The information was taken from the Ordnance Survey 1:25 000 (2½ in/1 mile) series, each point representing the main topographical feature in each ½ km square. The results have an accuracy of ±12·5 ft (3·81 m) as compared with the maps from which the information was extracted.

This work has pioneered the use of computer techniques in problems of radio network planning and frequency assignment, and will lead to better utilization of the radio spectrum and improved system planning. The JRC is producing further programs to predict the best arrangements of directional aerials to provide desired area coverage, and to produce radio path profiles and system design calculations for point-to-point links.

Further information on this subject can be obtained from the Secretary, The Joint Radio Committee of the Nationalized Power Industries, c/o 30 Millbank, London, SW1P 4RD.

U.D.C. 621.372.622

Minimum Conversion Loss and Input Match Conditions in the Broadband Mixer

Professor D. P. HOWSON,
D.Sc., C.Eng., F.I.E.E., F.I.E.R.E.†

List of Principal Symbols

A_L	conversion loss ratio
A_{opt}	optimum value of A_L
G	circuit conductance when $G_0 = G_{-2} = G$
G_0	circuit conductance at signal frequency
G_{-1}	circuit conductance at intermediate frequency
G_{-2}	circuit conductance at image frequency
G_{in}	input conductance
G_{out}	output conductance
$g(t)$	time-varying conductance
g_0	average value of $g(t)$
g_n, g_1, g_2	coefficients of $n\omega_p$, ω_p, $2\omega_p$ in Fourier expansion of $g(t)$
I_0	source current at signal frequency
I_{-1}	substitute source current at intermediate frequency
k	constant defining g_2, see equation (26)
R_D	forward resistance in diode model (Fig. 4)
R_R	reverse resistance in diode model (Fig. 4)
s	fraction of cycle $g(t)$ low conductance, bilinear model
t_d	diode noise temperature ratio
V_B	bias voltage
V_p	local-oscillator voltage
v_0	diode voltage component at signal frequency
v_{-1}	diode voltage component at intermediate frequency
v_{-2}	diode voltage component at image frequency
α	G_{-1}/G
α_{opt}	α for A_{opt}
β	$g_1^2/[g_0(g_0+g_2)]$
γ	$\sqrt{\{(2-k)/\delta-1\}}$
δ	constant defining g_1, g_2, see equation (26)
ε	constant defining G, see equation (30)
λ	$(\gamma-k+2)$
ω_p	local-oscillator frequency (angular)
ω_q	signal frequency (angular)
$(\omega_p-\omega_q)$	intermediate frequency (angular)
$(2\omega_p-\omega_q)$	image frequency (angular)

SUMMARY

The two approaches to the optimization of broadband mixer loss are compared and contrasted. Using a recently-developed model of the diode and local-oscillator circuit, calculations are made to show the difference in the loss in each case, and the mismatch required at the input to achieve minimum loss. It is concluded that the design procedure in which the mixer is matched at the input is practical, and that designing for minimum loss invoking mismatch is not to be recommended.

† *Postgraduate School of Electrical and Electronic Engineering, University of Bradford, Bradford 7, Yorkshire.*

1. Introduction

In the literature[1-11] on mixers, two approaches to the optimization of the broadband mixer are to be found. In one of these[1] the circuit is designed to be conjugately matched to the source and load, following which the resultant loss is calculated. The other approach[3] optimizes the circuit for minimum loss, following which the source and load impedances or admittances are calculated. It is not generally known that these optimization procedures are essentially incompatible for this particular circuit, although they lead to the same result for narrow-band mixers, i.e. those in which the image-frequency components are open- or short-circuited. The object of the present paper is to compare and contrast the two procedures, and to supplement the analytic

results with computer calculations using a recently developed model for the Schottky barrier diode.[7] Throughout the paper it is assumed that the diode can be treated as a time-varying conductance. The calculations are based on transducer gain formulae, available gain calculations leading to the same results except that the load impedance is not specified in the optimization procedures.

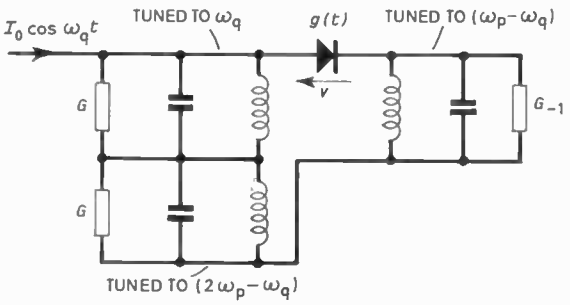


Fig. 1. Broadband series mixer circuit.

2. Analysis

Figure 1 shows a general 'broadband' series mixer circuit, the main characteristics of which are: a tuned i.f. termination; and a tuned signal termination which sustains the signal voltage (v_0) at a frequency ω_q, and the image frequency voltage (v_{-2}) at a frequency $(2\omega_p - \omega_q)$ and is therefore shown as two separate tuned circuits. At both these frequencies the terminating admittance is assumed to be G, and it is this broadband characteristic of the input circuit that gives the mixer its name. The mixer diode may be represented as a conductance periodically varying at the local-oscillator frequency, ω_p

$$g(t) = g_0 + 2\sum_{n=1}^{\infty} g_n \cos n\omega_p t \qquad \text{.....(1)}$$

if the local-oscillator completely controls the variation in the diode conductance, and reactive effects are negligible. Noting that the diode voltage is

$$v = v_0 \cos \omega_q t + v_{-1} \cos (\omega_q - \omega_p)t + v_{-2}(\omega_q - 2\omega_p)t \qquad (2)$$

a matrix equation for the circuit may be derived[4]

$$\begin{bmatrix} I_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} (G+g_0) & g_1 & g_2 \\ g_1 & (G_{-1}+g_0) & g_1 \\ g_2 & g_1 & (G+g_0) \end{bmatrix} \begin{bmatrix} v_0 \\ v_{-1} \\ v_{-2} \end{bmatrix} \qquad (3)$$

if it is assumed that the tuned circuits are ideal, in other words, have infinite admittance to all unwanted products.

Then it can be seen that

$$\frac{v_{-1}}{I_0} = \frac{g_1(G+g_0-g_2)}{(G+g_0)[(G_{-1}+g_0)(G+g_0)-g_1^2]- -g_1^2[G+g_0-g_2]+g_2[g_1^2-(G_{-1}+g_0)g_2]} \qquad \text{.....(4)}$$

Since the conversion loss is defined as

$$L = 10 \log_{10} A_L^2 = 10 \log_{10} \left\{ \frac{\text{available input power}}{\text{output power}} \right\} \qquad \text{.....(5)}$$

where

$$A_L = \frac{I_0}{2v_{-1}\sqrt{GG_{-1}}} \qquad \text{.....(6)}$$

from equation (4)

$$A_L = \frac{(G_{-1}+g_0)[G+g_0+g_2]-2g_1^2}{2g_1\sqrt{GG_{-1}}} \qquad \text{.....(7)}$$

If $G_{-1} = \alpha G$

$$A_L = \frac{(\alpha G+g_0)(G+g_0+g_2)-2g_1^2}{2g_1 G\sqrt{\alpha}} \qquad \text{.....(8)}$$

Then

$$\frac{dA_L}{d\alpha} = 0$$

when

$$\alpha^{\frac{1}{2}}.G(G+g_0+g_2) - -\tfrac{1}{2}\alpha^{-\frac{1}{2}}[(\alpha G+g_0)(G+g_0+g_2)-2g_1^2] = 0 \qquad \text{...(9)}$$

$$\alpha G(G+g_0+g_2) = g_0(G+g_0+g_2)-2g_1^2 \qquad \text{.....(10)}$$

and

$$A_L^2 = \frac{(G+g_0+g_2)[g_0(G+g_0+g_2)-2g_1^2]}{g_1^2 G} \qquad \text{...(11)}$$

$$\frac{dA_L^2}{dG} = 0$$

when

$$(G+g_0+g_2)[g_0(G+g_0+g_2)-2g_1^2] - -G[2g_0(G+g_0+g_2)-2g_1^2] = 0 \qquad \text{...(12)}$$

$$G^2 = [g_0(g_0+g_2)-2g_1^2]\frac{g_0+g_2}{g_0} \qquad \text{.....(13)}$$

This leads to

$$\alpha_{opt} = \frac{g_0}{g_0+g_2} \qquad \text{.....(14)}$$

and

$$A_{opt}^2 = \frac{1}{\beta}[1+\sqrt{1-2\beta}]^2 \qquad \text{.....(15)}$$

where

$$\beta = \frac{g_1^2}{g_0(g_0+g_2)}$$

This is the same result as that quoted by Strum[3] for minimum *available* conversion loss, and gives a minimum 3 dB loss when $\beta \to 0{\cdot}5$. β cannot be greater than 0·5 with a diode the characteristic of which has no region of negative resistance.[12]

3. Input and Output Admittances

Equation (3) will represent either the broad or narrow-band mixer if the termination at the image frequency is given the general value G_{-2}. The input admittance can then be calculated, since

$$\frac{I_0}{v_0} = G_{in} + G \qquad \text{.....(16)}$$

This gives

$$G_{in} = g_0 - \frac{g_2^2(G_{-1}+g_0)+g_1^2(G_{-2}+g_0)-2g_1^2g_2}{(G_{-1}+g_0)(G_{-2}+g_0)-g_1^2} \quad (17)$$

The output admittance is calculated by substituting a generator $I_{-1}\cos(\omega_q-\omega_p)t$ at the i.f. port for the signal generator, replacing the latter by an open circuit. Then

$$\frac{I_{-1}}{v_{-1}} = G_{out}+G_{-1} \quad \text{.....(18)}$$

gives

$$G_{out} = g_0 - \frac{g_1^2(G_{-2}+g_0)+g_1^2(G+g_0)-2g_1^2g_2}{(G+g_0)(G_{-2}+g_0)-g_2^2} \quad (19)$$

For the case of the 'broadband' mixer, equations (17) and (19) reduce to

$$G_{in} = g_0 - \frac{g_2^2(\alpha G+g_0)+g_1^2(G+g_0)-2g_1^2g_2}{(\alpha G+g_0)(G+g_0)-g_1^2} \quad (20)$$

$$G_{out} = g_0 - \frac{2g_1^2}{G+g_0+g_2} \quad \text{.....(21)}$$

Substituting into the latter from equation (13)

$$G_{out} = \frac{g_0}{g_0+g_2}\,G = \alpha_{opt}\,G = G_{-1} \quad \text{.....(22)}$$

so that the circuit is matched at the i.f. port under optimum conditions.

Examination of equation (20) confirms the general conclusion reached elsewhere,[5] that in a broadband mixer there is *not* a conjugate match at the input port when the circuit is optimized for minimum loss for given values of g_0, g_1 and g_2. However, it is possible that for one particular set of values of g_0, g_1 and g_2 minimum loss and input match may coincide. To obtain a match at the input port α may be controlled from equation (8), and G_{in} can be set equal to G in equation (20). This leads to the following cubic equation:[6]

$$g_0G^3+g_1^2(\beta^{-1}-1)G[G-g_0+g_2]- -g_1^2(\beta^{-1}-2)(g_0^2-g_2^2) = 0 \quad \text{...(23)}$$

The loss under these conditions can be calculated by substituting this value of G in equation (11). Although this will usually be greater than the minimum loss there may be conditions as just noted when they become equal. It is probably the loss under matched conditions that is of most interest to mixer designers, since this is a practical and easily recognizable condition. It is nevertheless important to realize that this loss may be greater than the minimum achievable, and computed results will be referred to a little later to estimate the magnitude of this extra loss.

An approximate analysis follows which is suitable for good quality diodes which conduct significantly for only a small fraction of the local oscillator cycle.

For a passive time-varying conductance

$$g_0 \geqslant |g_1|, |g_2| \quad \text{.....(24)}$$

and for a near-impulsive conductance waveform

$$g_0 \simeq g_1 \simeq g_2 \quad \text{.....(25)}$$

Take

$$g_1 = (1-\delta)g_o, g_2 = (1-2k\delta)g_o \quad (26)$$

for

$$\delta \ll 1, \qquad k < 2-\delta$$

as an approximation to this ideal condition. Examination of computer calculations for both the bilinear and the exponential diode model[7] confirms that these conditions match the computed results, usually with k between 1·75 and 1·97. The value of k cannot be equal to or greater than $(2-\delta)$ since this violates the condition for the diode to be a passive resistance.[12] It is advantageous for k to be as close to 2 as possible since then β is a maximum for a given δ, minimizing the loss, the expression being

$$\beta \simeq 0{\cdot}5\{1-(2-k)\delta\}$$

The ratio of i.f. conductance to source conductance

$$G_{-1}/G \simeq 0{\cdot}5(1+k\delta)$$

Solution of equation (13) for the optimum terminating conductance then gives

$$G_{opt} = 2\gamma\delta g_0 \quad \text{.....(27)}$$

where

$$\gamma = \sqrt{\frac{2-k}{\delta}-1}$$

and this leads to a value of the input conductance, from equation (17), of

$$G_{in}/G_{opt} = \frac{A}{B\gamma} \quad \text{.....(28)}$$

where

$$A = (\gamma^2+4\gamma+\gamma k+8k-4k^2)$$

and

$$B = (3\gamma+4-k)$$

The loss of the mixer is a minimum, for a particular diode, at a certain value of s, or V_B/V_P (see ref. 7), as below this value it is found that k decreases, finally approaching 0·5, and at the same time δ increases (see Appendix 1 for the form of variation). For these conditions γ is large and equation (28) reduces in the limit to

$$G_{in}/G_{opt} = 0{\cdot}33$$

Figure 3 shows that this figure is approached for values of $|V_B/V_P|$ approaching unity. For values of s or $|V_B/V_P|$ around and below the optimum, however, usually $\gamma \simeq 1$ and G_{in}/G_{opt} is greater than unity.

The minimum loss, from equation (11) is

$$A_L^2 \simeq \frac{2\lambda}{\gamma}\left[1+\left(\lambda-\frac{1}{\lambda}\right)\delta\right] \quad \text{.....(29)}$$

where

$$\lambda = (\gamma-k+2)$$

For the limiting case of an impulsive conductance waveform, examination of equation (23) shows that the mixer cannot be matched at both input and output. However, over a range of values of conductance coefficients of practical interest, matching is possible, and

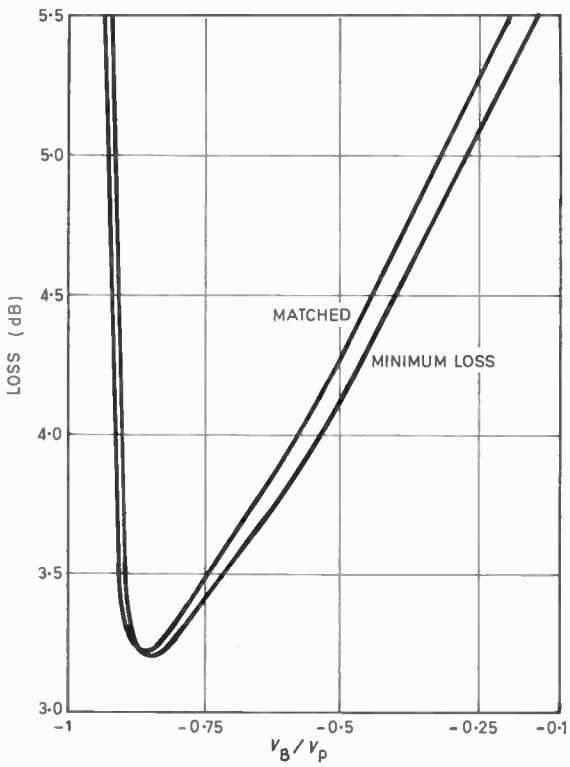


Fig. 2. Mixer loss for minimum loss and matched conditions as function of ratio of bias to l.o. voltage.

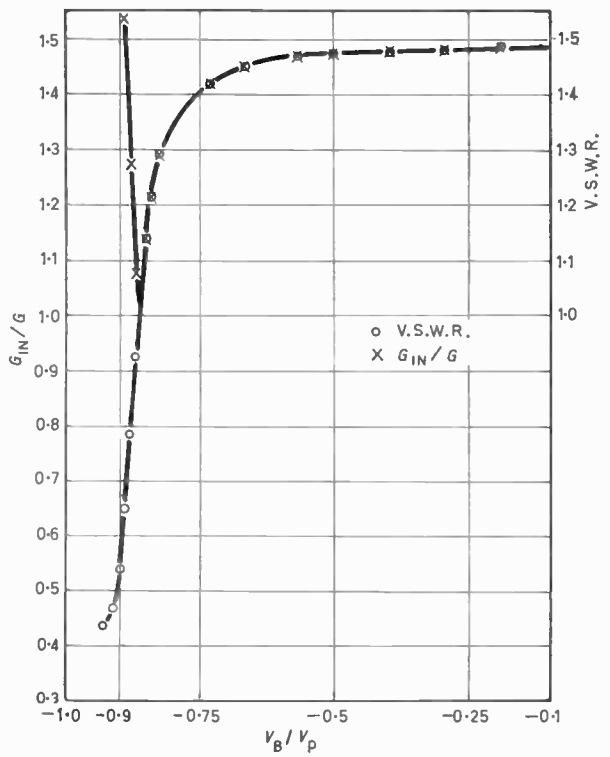


Fig. 3. Input conductance and v.s.w.r. at input of mixer for minimum loss as function of drive.

substitution of equation (26) into equation (23) gives as a rough approximation

$$G \simeq 2\varepsilon\delta g_0 \qquad \text{.....(30)}$$

where

$$2\varepsilon = k\left(1+\sqrt{\frac{8}{k}(2-k)+1}\right)$$

Note that if k is close to 2, this further simplifies to $G \simeq 2k\delta g_0$. The loss can be calculated from equation (29) by substituting ε for γ in the equation and the expression for λ.

If δ is small, the conductance waveform is impulsive in character, and may require a high-quality diode for its implementation. If the conductance variation is assumed to be two-state for example, and the fraction of the cycle for which it is low conductance is denoted by s, $\delta \simeq 0{\cdot}1$ implies $s \simeq 0{\cdot}88$. Alternatively, a more accurate modelling of the diode using an exponential model for the junction[7] would give a rather lower value, say $s \simeq 0{\cdot}8$. Barber[11] has estimated the 'cut-off' frequency of the diode, effectively in terms of the maximum achievable value of s, and in these two cases 'cut-off' frequencies of 650 GHz and 500 GHz respectively would be necessary. This should be achievable with modern Schottky-barrier diodes.

Figure 2 shows the loss of the mixer under minimum loss and matched conditions, as a function of the ratio of bias to local-oscillator voltage. Figure 3 shows input conductance and the v.s.w.r. at the input for minimum loss with various drive conditions. For both these calculations the diode model, assumed to be in an untuned local-oscillator circuit, consisted of an exponential junction in series and parallel with linear resistors[7] (see Fig. 4). The diode parameters used were taken from manufacturers' data for typical Schottky-barrier diodes (Hewlett-Packard type 2800) and were:

$$R_D = 11\ \Omega, \quad \alpha_1 = 40\ V^{-1} \quad \text{and} \quad I_S = 1\times10^{-9}\ A,$$

where the exponential relationship between junction current and voltage is given by $I_D = I_s\{\exp(\alpha_1 V_j)-1\}$.

A reverse resistance R_R of $1\times10^5\ \Omega$ was assumed and the local oscillator voltage was taken to be 3 V peak supplied by a source of output resistance 50 Ω. Virtually no difference is appreciable at this drive level with output resistances of between 5 Ω and 5 kΩ.

Several conclusions can be drawn from Figs. 2 and 3. It is clear that the difference in the two loss calculations is small, being less than 0·1 dB over a wide range of bias conditions around the optimum. More detailed examination of the lowest loss achievable with this diode

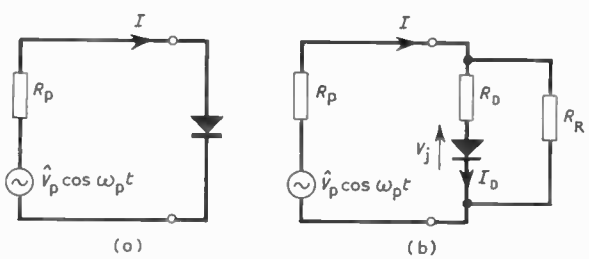


Fig. 4. (a) Elementary untuned local oscillator circuit. (b) Diode model.

and local-oscillator circuit reveals that the mixer is not conjugately matched under these conditions. However, the point at which minimum loss coincides with conjugate matching is so close to the point of lowest loss as to be indistinguishable on Fig. 2, so that for all practical purposes they may be taken to coincide for this mixer.

Similar calculations were made for the dual broadband shunt mixer, in which the same diode and local-oscillator circuit were incorporated in a dual structure, the signal and i.f. filters now being series rather than parallel tuned circuits. Similar results were achieved, with the value of lowest loss being 3·18 dB, a slight improvement (0·04 dB) over the more easily realizable circuit with which the paper is primarily concerned. This improvement is in line with that noted for the narrowband mixer elsewhere.[7]

The input conductance results in Fig. 3 bear out the earlier approximate analysis, and show a remarkable transition as the circuit approaches lowest loss. The results suggest that the design of the mixer for an input mismatch in the hopes of lowering the loss would be quite impracticable with this circuit in this region, because of the extreme sensitivity of the input conductance to the ratio of bias to local oscillator voltage. Furthermore, in the region where the input conductance is not sensitive to changes in V_B/V_P, the v.s.w.r. is unacceptably poor. It therefore seems that, if the circuit model chosen is representative of modern mixers, it is very fortunate that the increased loss over the minimum that occurs when the mixer is matched at input and output is so small.

If the diode noise temperature ratio, t_d is approximately unity,[8] then, for a mixer loss of over 2 dB and an i.f. amplifier noise figure greater than 1 dB, minimum conversion loss will coincide with minimum noise figure.[9] Even if it is necessary to take into account the diode capacitance in calculating conversion loss it has been shown experimentally that minimum noise figure for the mixer usually coincides with minimum conversion loss calculated using the resistive diode model, i.e. neglecting the capacitance effect.[10]

4. Acknowledgment

The author gratefully acknowledges the help gained in numerous discussions with Dr. J. G. Gardiner and the help in preparing the computer programs from Mrs. J. Butland.

5. References

1. Torrey, H. C. and Whitmer, C. A., 'Crystal Rectifiers' (McGraw Hill, New York, 1948).
2. Macpherson, A. C., 'An analysis of the diode mixer consisting of nonlinear capacitance and conductance and ohmic spreading resistance', *Trans. Inst. Radio Engrs on Microwave Theory and Technique*, MTT-5, p. 43, January 1957.
3. Strum, P. D., 'Some aspects of mixer crystal performance', *Proc. Inst. Radio Engrs*, 41, p. 875, July 1953.
4. Howson, D. P., 'Minimum loss of a general broad-band mixer', *Electronic Letters*, 5, p. 123, 20th March 1969.
5. Howson, D. P., 'Maximum power transfer in parametric circuits', Summer School on Circuit Theory, Talé, Czechoslovakia, 1971.
6. O'Neill, H. J., 'Image-frequency effects in a microwave crystal mixer', *Proc. Instn Elect. Engrs*, 112, p. 2019, November 1965.
7. Glover, K. J., Gardiner, J. G. and Howson, D. P., 'Double-tuned modulator calculations using an improved diode model'. *The Radio and Electronic Engineer*, 42, pp. 37–44, January 1972.
8. Hewlett Packard Application Note No. 907.
9. Vugrinic, J., 'Input impedance considerations for a . . . mixer followed by a . . . preamplifier', read at Summer School on Circuit Theory, Prague, 1968.
10. Liechti, C. I., 'Down-converters using Schottky-barrier diodes', *Trans. I.E.E.E. on Electron Devices*, ED-17, p. 975, November 1970.
11. Barber, M. R., 'Noise figure and conversion loss of the Schottky-barrier mixer diode', *Trans. I.E.E.E.*, MTT-15, p. 629, November 1967.
12. Howson, D. P., 'Passivity constraints on mixer diode coefficients', *Electronics Letters*, 8, p. 4, 13th January 1972.
13. Peterson, L. C. and Llewelyn, F. B., 'The performance and measurement of mixers in terms of linear network theory', *Proc. Inst. Radio Engrs*, 33, p. 458, July 1945.

6. Appendix 1

The Bilinear Approximation for Diode Conductance

It has been shown[7] that this elementary model, which assumes that the diode conductance always takes one of two values, gives a reasonable approximation to the actual values of diode conductance coefficients over most of the range. It is used here to show the *form* of variation that the parameters δ and k take for varying values of s, and computed results using more exact diode models show these to be typical. The Fourier expansion for $g(t)$ using this approximation is

$$g(t) = g_b + (1-s)(g_f - g_b) + \frac{2}{\pi}(g_f - g_b)\sum_{n=1}^{\infty}\frac{\sin n\pi(1-s)}{n}\cos n\,\omega_p\, t \quad (31)$$

where g_f is the forward-biased conductance and g_b the reverse-biased conductance.

Then, from equations (1) and (31)

$$\frac{g_1}{g_0} = \frac{\sin \pi(1-s)}{\pi(1-s)} \cdot \frac{g_f - g_b}{g_f + \left(\frac{s}{1-s}\right) g_b}$$

If $g_f = n^2 g_b$, n denoting the quality of the diode,

$$\frac{g_1}{g_0} = \frac{\sin \pi(1-s)}{\pi(1-s)} \cdot \frac{1 - 1/n^2}{1 + \frac{s}{(1-s)n^2}}$$

For $(1-s) < 0{\cdot}1$, and $n^2 > 10^3$

$$\frac{g_1}{g_0} \simeq \frac{1-\pi^2(1-s)^2}{3} - \frac{1}{(1-s)n^2} \quad \ldots\ldots(32)$$

Similarly,

$$\frac{g_2}{g_0} \simeq \frac{1-4\pi^2(1-s)^2}{3} - \frac{1}{(1-s)n^2}$$

Above the optimum value of $(1-s)$ the second terms on the right-hand side are much larger than the third. Hence the result quoted in equation (26), with δ increasing as $(1-s)$ increases, and k close to 2. Below

the optimum value of $(1-s)$ a region is reached where the third terms are much larger than the second. Then δ increases inversely with $(1-s)$ and k approaches 0·5. This leads to the sharp increase of loss with increasing s noted below the minimum value, and the change of value of G_{in}/G_{out} already noted. If the third term were not present the loss would continue to decrease as s increases. The relative importance of the third term is clearly inversely proportional to n^2, explaining the well-known result that the better the diode quality the larger the value of s for minimum loss.

7. Appendix 2

Peterson and Llewellyn's Analysis

It should be noted that other authors[13] have implied that the broadband mixer is image matched for minimum loss, which is clearly not the case, as is shown in Fig. 3 and the preceding text. The error of reasoning in the work of Peterson and Llewellyn follows from their method of analysis. They reason that, since $v_0 = v_{-2}$ when the mixer is driven by a current at the i.f. port, $I_{-1}\cos(\omega_q-\omega_p)t$, then the analysis can be simplified for this type of operation by replacing v_{-2} by v_0 in equation (3). This leads to the matrix equation for an equivalent two-port:

$$\begin{bmatrix} 0 \\ I_{-1} \end{bmatrix} = \begin{bmatrix} g_0+g_2+G & g_1 \\ 2g_1 & g_2+\alpha G \end{bmatrix} \begin{bmatrix} v_0 \\ v_{-1} \end{bmatrix} \quad \ldots(33)$$

the third equation in (3) being discarded as redundant. General theory[5] shows that this circuit must be conjugately matched for minimum loss, and the correct values of α, G and A_L follow. By reciprocity[4] the loss is the same from ω_q to $(\omega_q-\omega_p)$ as from $(\omega_q-\omega_p)$ to ω_q, so that the calculated loss is the required minimum loss of the mixer. Moreover, since the port terminations are real, the circuit is also image matched at input and output.

The fallacy is that the 'equivalent' circuit is not an exact equivalent of the broadband mixer, and in particular the input admittance of the 'equivalent' circuit is

$$G'_{in} = g_0 + g_2 - \frac{2g_1^2}{g_0+\alpha G} \quad \ldots\ldots(34)$$

which is not the same as that of the broadband mixer, given by equation (20). Thus although the 'equivalent' circuit is image-matched for minimum loss, it does *not* follow that the broadband mixer is image-matched for minimum loss. As pointed out in the main text, there is, in the general case, a mismatch at the input.

Manuscript first received by the Institution on 19th October 1971 and in revised form on 6th January 1972. (Paper No. 1450/CC128).

STANDARD FREQUENCY TRANSMISSIONS—April 1972

(Communication from the National Physical Laboratory)

April 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		April 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	+0·1	0	0	625	648.5	17	—	0	+0·1	—	648·9
2	0	+0·1	0	625	648·0	18	—	+0·1	+0·1	—	648·4
3	0	+0·1	0	625	647·7	19	—	+0·1	+0·1	—	647·3
4	0	−0·1	0	625	649·1	20	—	+0·1	+0·1	—	665·8
5	0	0	−0·1	625	649·2	21	—	+0·1	+0·1	—	663·8
6	0	0	0	625	649·2	22	—	+0·1	0	—	663·0
7	0	0	0	625	649·5	23	—	+0·1	+0·1	—	662·0
8	0	0	0	625	649·5	24	—	+0·2	+0·1	—	656·9
9	0	0	0	625	649·2	25	—	+0·1	+0·1	—	655·9
10	—	+0·1	0	—	648·3	26	—	+0·1	0	—	654·6
11	—	+0·1	−0·1	—	647·0	27	—	0	+0·1	—	654·6
12	—	−0·1	0	—	651·4	28	—	+0·1	+0·1	—	650·5
13	—	0	+0·1	—	651·4	29	—	−0·1	0	—	651·8
14	—	+0·1	0	—	650·6	30	—	+0·2	+0·1	—	649·7
15	—	0	+0·1	—	650·4						
16	—	+0·1	+0·1	—	649·3						

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

* Relative to UTC Scale; $(UTC_{NPL} - Station) = +500$ at 1500 UT 31st December 1968.

† Relative to AT Scale; $(AT_{NPL} - Station) = +468{\cdot}6$ at 1500 UT 31st December 1968.

The GBR Transmitter at Rugby is off the air for maintenance until the end of July.

U.D.C. 535.375.54: 621.373.826

A Tunable Spin-flip Magneto-Raman Infra-red Laser

R. L. ALLWOOD, B.Sc.†

R. B. DENNIS, B.Sc., Ph.D., M.Inst.P.,†

W. J. FIRTH, B.Sc.†

Professor S. D. SMITH, B.Sc., Ph.D., D.Sc., F.Inst.P.†

B. S. WHERRETT, B.Sc., Ph.D.†

and

R. A. WOOD, M.A., D.Phil.†

Based on a paper presented at the I.E.R.E. Conference on Infra-red Techniques held in Reading on 21st to 23rd September, 1971.

SUMMARY

The observation of intense, magnetic-field tunable, spin-flip Raman scattering in indium antimonide, pumped at two wavelengths, namely, 10·6 μm and 5·3 μm is reported. The observed tuning range is 9 μm to 14 μm, and 5·4 μm to 5·7 μm; the physical processes and experimental techniques are described, and possibilities for extension of the tuning range are discussed. Some future applications of such an intense infra-red source are mentioned.

† *Physics Department, Heriot-Watt University, Riccarton, Currie, Midlothian, U.K.*

1. Introduction

In this paper we review some aspects of the tunable stimulated Raman scattering from the conduction electrons in indium antimonide, first observed in 1970.[1-3] A brief description of the spin-flip Raman scattering process is followed by an explanation of the experimental set-up used and by a detailed discussion of the results obtained to date. A brief outline of the major aims of future experimental programme is presented and finally mention is made of the many possible applications of this powerful, highly monochromatic, tunable laser source.

2. Spin-flip Raman Scattering

Raman scattering is an inelastic scattering process in which photons of energy $\hbar\omega_0$ interact with a scattering medium to produce photons of a different frequency ω_1 given by

$$\hbar\omega_1 = \hbar\omega_0 \pm E'$$

where E' is the energy of the accompanying transition of the medium. Scattering to a lower frequency is termed Stokes scattering ($\hbar\omega_s = \hbar\omega_0 - E'$), that to a higher frequency, anti-Stokes scattering ($\hbar\omega_{as} = \hbar\omega_0 + E'$). Higher order processes may also take place, among which are double-Stokes scattering processes, in which photons of energy $\hbar\omega_{2s} = \hbar\omega_0 - 2E'$ are produced. The symbols ω_{as} and ω_{2s} refer to the frequencies of anti-Stokes and double-Stokes scattered radiation respectively.

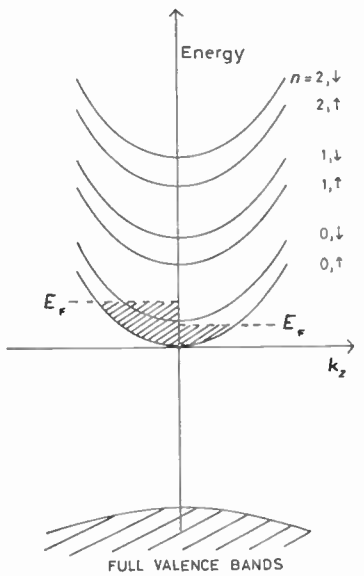


Fig. 1. Magnetic conduction band of InSb. Left, high carrier concentration leading to blocking effects. Right, $n=0, \downarrow$ level unoccupied, thus giving maximum scattering efficiency.

Of particular interest is magneto-Raman scattering from conduction electrons in n-type indium antimonide (InSb). Figure 1 shows the band structure of InSb in a uniform d.c. magnetic field. The scattering process reported in this paper is associated with spin-reversal transitions of the electrons between the spin-up $(0, \uparrow)$ and spin-down $(0, \downarrow)$ levels of the conduction band. The energy difference between these spin states is given by $|g^*|\beta B$, where g^* is the conduction band g-factor, β the Bohr magneton, and B the applied magnetic field.

Indium antimonide has an exceptionally large g-factor (~ -45), and so is particularly suitable for spin-flip scattering, having a large tuning range per unit magnetic field.

Spin-flip Raman scattering was predicted by Yafet[4] in 1966, who considered a simplified model with a single conduction electron, and it was first observed as 'spontaneous' scattering in 1967.[5] Raman scattering can, however, become stimulated, the presence of Stokes photons inducing further scattering, leading to laser action at the Stokes frequency. Measurements of the spin-flip Raman cross-section led to an estimated stimulation threshold in InSb which was low enough to be attainable using available CO_2 lasers to pump at 10·6 μm.

Stimulation was not, however, achieved until 1970[1-3] after Wherrett and Harper[6] had extended the theory to finite electron concentrations, incorporating the effect of Fermi statistics, which leads to strong magnetic field and concentration dependence of the scattering efficiency. Figure 1 shows the energy-wavevector (in the z-direction, k_z) for the conduction electrons of InSb in the presence of an applied magnetic field. Since the momentum transfer from an incident photon to the electron is small, the spin-flip transition, from the $n=0, \uparrow$ to the $n=0, \downarrow$ state, is nearly vertical in this E vs. k_z diagram. Any population of the upper level ($n=0, \downarrow$) as shown by Fig. 1, thus has a blocking effect, due to the Pauli exclusion principle, and so reduces the overall scattering efficiency. Maximum scattering efficiency is thus only obtained when the upper spin level is completely empty. This occurs only when the magnetic field is greater than a certain minimum value (the 'quantum limit'). This value increases with carrier concentration, and sets an upper limit to the carrier concentration for which stimulation can be achieved, the optimum concentration being about 10^{16} cm^{-3}.

For a specimen of length l the threshold condition for Raman laser oscillation to occur in the collinear direction (parallel to the pump beam) is $R\exp(g-\alpha)\,l > 1$, where R is the sample reflectivity, α the absorption coefficient and g the Raman gain given by

$$g = \frac{16\pi^2 c^3 S I}{\hbar\omega_s^3 n_s^2 \Gamma}$$

Here S is the overall scattering efficiency (cross-section per electron × concentration of 'unblocked' electrons), I the incident pump radiation intensity, ω_s the Stokes frequency (rad s^{-1}), Γ the *spontaneous* Stokes linewidth and n_s is the refractive index at the Stokes frequency.

The Raman efficiency S is frequency-dependent and shows strong enhancement when the pump photon energy ($\hbar\omega_0$) is close to the (magnetic field-dependent) InSb band gap. The resonance in S actually occurs when the pump photon energy is equal to the energy difference between the valence band and the *higher* of the two spin levels ($E_g + |g^*|\beta B$). Comparison of the gain and loss factors at 4·0 T for 10·6 μm and 5·3 μm pump wavelengths, indicates that the threshold for the latter should be lower by a factor of approximately 10, primarily due to this resonant enhancement. Experimentally the ratio of the thresholds (3×10^5 Wcm^{-2} and 10^4 Wcm^{-2}), is in reasonable agreement with this figure.

3. Experimental Details

In these experiments, plane polarized radiation at 10·6 μm was obtained from a Q-switched CO_2 laser. The laser consisted of a flowing gas discharge tube, length about 2 m, the cavity being formed by a semi-transmitting flat germanium output mirror and a 10 m focal length, concave, gold plated reflector. Q-switching was by means of a rotating plane mirror inside the cavity. Also within the cavity was a diffraction grating with angular adjustment for laser line selection, and an aperture to ensure single lowest order transverse mode operation. The laser delivered pulses of plane polarized radiation having peak intensity of nearly 3 kW, of approximate duration of 300 ns and with a repetition rate of 100 Hz.

The radiation was focused, by means of a 200 mm focal length mirror, into a specimen of indium antimonide mounted on a cold finger in the bore of a superconducting solenoid. Specimens used were typically of dimensions $7\times2\times2$ mm with their end faces polished flat and parallel to form the Raman laser cavity. The radiation travelled perpendicular to the magnetic field, with its electric vector parallel to the field. The output radiation in the collinear scattering direction was collected and focused on to the slit of a $f/5$ grating monochromator, used for wavelength analysis and partial rejection of the intense transmitted pump beam. The radiation emerging from the monochromator was monitored using a liquid helium-cooled Ge:Cu detector, after further 10·6 μm rejection by dielectric interference filters. The filters used were specifically designed for the purpose, having a transmission as low as 1% at 10·6 μm and as high as 70% at 10·9 μm. Signals were either displayed directly on to an oscilloscope or via a box-car integrator on to a chart recorder.

Intense magnetic field tunable scattered radiation is observed from InSb at frequencies ω consistent with the relations $\omega = \omega_0 + \omega_{sf}$, $\omega_0 - \omega_{sf}$ and $\omega_0 - 2\omega_{sf}$.[7] Here ω_0 is the pump angular frequency as before and ω_{sf} is the 'spin-flip frequency', $(1/\hbar)|g^*|\beta B$. For a peak pump power in the specimen of 1 kW focused to an area of about 10^{-3} cm^2, the peak scattered powers were observed to be: Stokes ($\omega_0-\omega_{sf}$) 10 W; anti-Stokes ($\omega_0+\omega_{sf}$) 0·1 W; double-Stokes ($\omega_0-2\omega_{sf}$) 0·02 W. The wave-number shifts of the experimentally observed radiation are plotted against applied magnetic field in Fig. 2. It is seen that, while intense Stokes and anti-Stokes radiations are both emitted from 3·0 T to 10·0 T, double-Stokes emission occurs only between 3·4 T and 6·5 T. For the stimulated Stokes radiation the lower magnetic field limit is due to the decrease in the spontaneous Stokes scattering efficiency at fields below the quantum limit.[6] The upper magnetic field limit is probably due to the increasing free carrier absorption coefficient, leading to the condition $R\exp(g-\alpha)\,l < 1$, for the maximum input intensity obtainable.

The production processes for the intense emissions at the anti-Stokes and double-Stokes frequencies have not

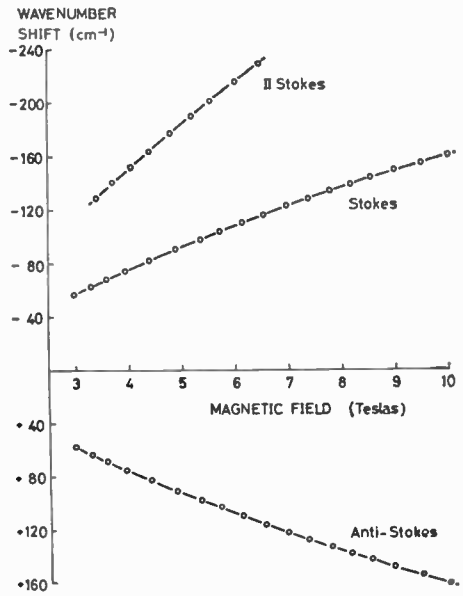


Fig. 2. Graphs showing wave-number shifts of experimentally observed radiations against applied magnetic field.

yet been identified with certainty. It is believed[8] that strong coupling of the Stokes and anti-Stokes waves leads to a 'mixing', so that the stimulated Stokes wave is 'followed' by a somewhat weaker anti-Stokes wave. This theory is supported by the observation that Stokes and anti-Stokes radiation have the same polarization (perpendicular to the pump polarization), the same magnetic field tuning range, and that the time development of the pulses is similar.

The mechanism of double-Stokes generation is still something of a mystery. Its polarization, plane, at nearly 30 deg to the 10·6 μm pump, seems to preclude the straightforward rescattering of Stokes radiation, but this would however explain the more restricted tuning range of 3·4 T to 6·5 T, as being a threshold effect.

For the 10·6 μm pump, the present tuning range has been shown to extend from 9 μm to 14 μm. The linewidth of the Stokes scattered radiation is known to be not greater than 0·03 cm^{-1}. The combination of high intensity and narrow linewidth lends itself to a convenient tunable, high power, coherent infra-red source.

Our group has recently extended the tuning range[9] to include the range 5 μm to 6 μm The CO_2 laser radiation was focused on to a crystal of tellurium, oriented for phase-matched second harmonic generation, producing nearly 20 W peak power of 5·3 μm radiation. Transmitted 10·6 μm radiation was reduced by a MgO filter, and the 5·3 μm radiation was re-collimated and focused on to the InSb specimen as before. Elimination of 10·6 μm radiation is desirable so as to prevent two photon absorption processes in the InSb specimen. The scattered radiation was analysed as before, again employing suitable interference filters and the same Ge:Cu detector.

The observed tuning range extended from 2·0 T to 6·4 T (corresponding to 5·4 μm and 5·7 μm). The peak intensity of the scattered radiation was very nearly 300 mW at 3·85 T, corresponding to a conversion efficiency of 1·5% in the specimen. It is the resonant enhancement of the scattering efficiency, discussed previously, that allows stimulation to occur for such low pump intensities.

Superimposed on the observed magnetic field dependence of this stimulated Stokes output (Fig. 3(b)) were sharp intensity minima at frequencies which correspond accurately to atmospheric water vapour absorption lines. Figure 3 shows a comparison between the observed output power as a function of magnetic field, i.e. wavenumber, and the saturated water vapour spectrum obtained by using a conventional infra-red grating spectrometer. In this crude illustration of this spectroscopic application of the spin-flip Raman laser, it can be seen that a resolving power at least equal to that of the conventional spectrometer (0·1 cm^{-1}) is obtained.

4. Future Work

One of the most exciting problems raised by this work is to develop a powerful tunable monochromatic source in the far infra-red region, i.e. beyond 25 μm. This may be best achieved by the difference frequency generation in a non-linear medium from two intense radiations of slightly different frequency. If one of these frequencies were tunable, then the generated radiation of lower energy would also be tunable. Mixing the pump radiation (10·6 μm) and the Stokes radiation (11 μm to 13 μm), would give an output tunable from 55 μm to 300 μm. Stimulated Stokes intensities as high as 1 kW have been reported,[10] and the use of phase matched difference frequency generation[11] should result in the production of at least 1 μW of monochromatic long wavelength radiation. Such powers would be large compared with conventional sources in this region of the spectrum.

The extension of the tuning range of the magneto-Raman laser below 55 μm wavelength is also of considerable interest. There are several approaches to this problem. The most straightforward solution is to have at hand a number of well-spaced pump laser frequencies so

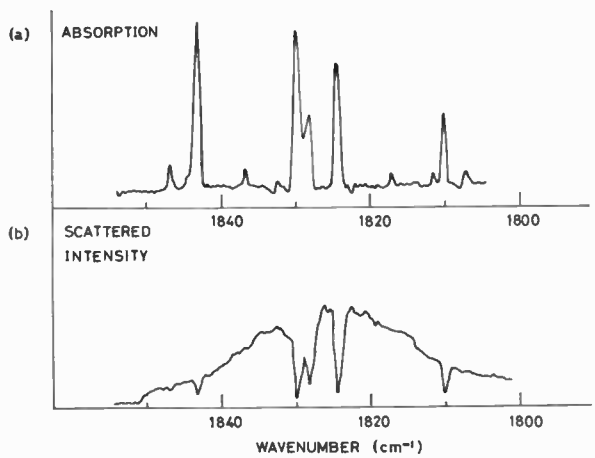


Fig. 3. Comparison between the observed output power as a function of magnetic field and the saturated water vapour spectrum obtained by using a conventional infra-red grating spectrometer.

chosen that the spin-flip tuning ranges overlap, giving complete coverage of the desired spectral region.

Another method is to use mixing techniques again, for example, producing the second-harmonic frequency of the Raman scattered radiation. Such techniques used at present would enable the production of radiation in the 4·5 μm to 7 μm wavelength range.

Lastly, there remains the need to investigate the scattering properties of other materials, in particular those having even larger g-factors than InSb. $Cd_xHg_{1-x}Te$, for example, may be particularly useful in this respect.

5. Applications

The spin-flip laser combines the tunability of conventional infra-red sources with the advantages of the laser, namely, high power, strong collimation and narrow linewidth. It thus improves many of the parameters of infra-red spectroscopy by several orders of magnitude, in particular a factor of over a million in power per unit frequency range.

The tunable laser will also be useful as a selective pump; in solid state physics it will, for example, be useful in the study of hot electron phenomena, while in chemistry it will find application as a selective catalyst. The tunable laser will also contribute to non-linear optics, enabling, for the first time, the study of the dispersion and resonance structure of non-linear susceptibilities.

In conclusion, this tunable infra-red laser, which has been developed, is a revolutionary new tool and should contribute widely to the advancement of science and technology.

6. References

1. Allwood, R. L., Smith, S. D., Devine, S. D., Wood, R. A., and Mellish, R. G., 'Tunable stimulated spin-flip magneto-Raman scattering with dielectric coated cavities,' *J. Phys. C*,, **3**, p. 186, 1970.
2. Patel, C. K. N. and Shaw, E. D., 'Tunable stimulated Raman scattering from conduction electrons in InSb', *Phys. Rev. Letts*, **24**, p. 451, 2nd March 1970.
3. Mooradian, A., Brueck, S. R. J., and Blum, F. A., 'Continuous stimulated spin-flip Raman scattering in InSb (using CO laser)', *Appl. Phys. Letts.*, **17**, p. 481, 1st December 1970.
4. Yafet, Y., 'Raman scattering by carriers in Landau levels', *Phys. Rev.*, **152**, p. 858, 9th December 1966.
5. Slusher, R. E., Patel, C. K. N., and Fleury, P. A., 'Inelastic light scattering from Landau-level electrons in semiconductors', *Phys. Rev. Letts*, **18**, p. 77, 16th January 1967.
6. Wherrett, B. S. and Harper, P. G., 'Concentration and magnetic-field dependance of spin-flip magneto-Raman scattering', *Phys. Rev.*, **183**, p. 692, 1969.
7. Allwood, R. L., Dennis, R. B., Smith, S. D., Wherrett, B. S., and Wood, R. A., 'Increased tuning range of intense magneto spin-flip scattering of 10·6 μm radiation in InSb', *J. Phys. C*, **4**, p. 163, 1st March 1971.
8. Bloembergen, N., 'Nonlinear Optics' (Benjamin, New York, 1965).
9. Allwood, R. L., Dennis, R. B., Mellish, R. G., Smith, S. D., Wherrett, B. S., and Wood, R. A., 'Spin-flip Raman scattering of harmonically generated 5·3 μm radiation', *J. Phys. C*, **4**, p. 126, 4th June 1971.
10. Aggarwal, R. L., Lax, B., Chase, C. E., Pidgeon, C. R., Limbert, D., and Brown, F., 'High intensity tunable InSb spin-flip Raman laser', *Appl. Phys. Letts.* **18**, p. 383, 1st May 1971.
11. Patel, C. K. N. and van Tran, N., 'Phase-matched nonlinear interaction between circularly polarized waves', *Appl. Phys. Letts*, **15**, p. 189, 15th September 1969.

Manuscript received by the Institution on 28th May 1971 and in final form on 16th February 1972. (Short Contribution No. 155/CC129).

Contributors to this issue

Dr. B. G. Evans obtained the degree of B.Sc. in electrical engineering at the University of Leeds in 1965. From 1965 to 1968 he carried out research into broadband microwave components and antennas at the University of Leeds, and obtained his Ph.D. in 1968. He is at present a Lecturer in the Department of Electrical Engineering Science at the University of Essex, where he is pursuing research into broadband antennas and propagation studies in digital microwave and satellite communication systems.

Dr. B. S. Wherrett obtained a B.Sc. degree in physics with first class honours at the University of Reading in 1966. He then proceeded to postgraduate work at the University and was awarded his Ph.D. in 1969. He held a Post Doctorate appointment at Reading for a year and then moved to Heriot-Watt University, where he is currently a Lecturer in the Physics Department.

Mr. W. J. Firth is a graduate of Edinburgh University, having obtained a B.Sc. degree with honours in mathematical physics in 1966. The following year he obtained the University's diploma in elementary particle physics. Mr. Firth was appointed to a Lectureship in the Physics Department of Heriot-Watt University in 1967. He joined the infra-red group under Professor Smith in November 1970.

Dr. R. A. Wood gained his first degree in physics at the University of Oxford in 1966. He then embarked on postgraduate studies at Oxford, being granted his D. Phil. in 1970. Dr. Wood currently holds a S.R.C. Research Fellowship in the Physics Department at Heriot-Watt University.

Dr. R. B. Dennis obtained his education at the University of Reading where he was awarded a first-class honours degree in physics in 1966 and a Ph.D. in 1969 for his magneto-optical investigations of the electronic and coupled electron-phonon properties in polar semiconductors. For a year he was employed as a Senior Research Officer at Reading University and he was then awarded a Science Research Council Fellowship tenable at Heriot-Watt University. During this time he spent a total of a year as a guest of the Physikalisches Institut der Universitat, Freiburg, West Germany, where he began work on non-linear optics, in particular harmonic generation and Raman scattering. Dr. Dennis now holds the post of Departmental Research Assistant with lecturing duties at Heriot-Watt University.

Mr. R. L. Allwood received a B.Sc. degree with first class honours in physics from the University of Reading in 1969. He then went to Heriot-Watt University as a post graduate research student and has recently completed his Ph.D. thesis. He has been closely involved with the research program of the Department and is a co-author of eight publications.

Mr. P. A. Ramsdale graduated with first-class honours in electrical engineering from the University of Manchester Institute of Science and Technology in 1967, and in the same year joined the Plessey Company where he worked on research and development of communication systems. In 1968 he went to the University of Birmingham to study for the degree of Ph.D., his thesis dealing with active aerials. In September 1971 he was appointed to a Lectureship in the Department of Electrical Engineering, Lanchester Polytechnic, Coventry.

Dr. T. S. M. Maclean who is a former Caird Scholar in Electricity, graduated from the University of Glasgow and received his doctorate from the University of Edinburgh. He works in the field of applied electromagnetism and has invented several radio aerials and done analytical and experimental work towards the elucidation of others. Several papers on this work have been published in the Institution's *Journal*. Since 1965 Dr. Maclean has held the established Senior Lectureship in Electromagnetism at the University of Birmingham, where he lectures on radiation and propagation, and microwave engineering. He also conducts a discussion group as part of the Departmental Broader Education programme.

Conference on Semiconductor Injection Lasers and their Applications

University of Cardiff
20th to 21st March 1972

The meeting was arranged jointly by the Department of Applied Mathematics and Mathematical Physics, University College, Cardiff, and the Department of Applied Physics, U.W.I.S.T., in association with the Institute of Physics, the Institution of Electrical Engineers, and the Institution of Electronic and Radio Engineers. The attendance at the meeting was over 60 and included representatives from France, Germany, and Switzerland. (Abstracts only of papers are available from the organizers, price 50p per set).

The background to the conference was introduced by the first session chairman, C. H. Gooch (S.E.R.L., Baldock), who recalled the historical progress of the injection laser over the decade since its announcement in 1962. Room-temperature threshold current densities have been reduced by a factor of over 100 to a present-day best of less than 1000A/cm^2, the most significant improvements occurring with the introduction of the $Al_xGa_{1-x}As$-GaAs heterostructure lasers in 1969. A review of current state-of-the-art devices was given by P. R. Selway (S.T.L., Harlow) who included details of the new mesa-structure and l.o.c. (large optical cavity) lasers. The improved theoretical understanding of the injection laser achieved in recent years has led to developments of this kind, which attempt to confine both the recombining carriers and the laser radiation field to well-defined regions of the device. Hence the use of a mesa-structure, for example, confines the active lasing region to a width of 10–20 μm and a thickness of 0·2–0·5 μm resulting in threshold currents as low as 100 mA.

A subject which received much attention at the meeting was that of laser degradation—both catastrophic and gradual. The former phenomenon occurs in the form of damage to the laser facets due to high optical flux densities; 'burn-off' power densities of the order of 100 Wmm^{-1} were the best values for single heterostructure lasers of 1·6 μm active region thickness mentioned by Mr. Selway. It is hoped that the new l.o.c. and l.g.r. (localization of gain to certain regions of the resonator) lasers, which attempt to reduce these flux densities, will lead to improvements in this field. A contribution by R. Keller *et al.* (Institute of Applied Physics, Berne) dealt with mirror distortions as a result of the elongation of the laser during a current pulse. This investigation made use of an ingenious interferometric measurement technique (involving a He-Ne laser) which recorded elongations of between 4 and 20 nm with good accuracy. These measured values corresponded to a heating of 1 deg–6 deg, reasonably agreeing with theory.

The topic of non-catastrophic degradation was dealt with in a contributed paper by D. H. Newman and S. Ritchie (P.O.R.D., Martlesham). Typical double heterostructure laser lifetimes are at present of the order of 100 hours 'on'-time, and this forms the main obstacle to the use of the device in long-range optical communications systems. The basic mechanisms responsible for degradation are as yet not understood, and several of the proposed explanations of this effect were reviewed. Dr. Newman's results indicated that degradation is associated with the accumulation of defects, at or near the active region of the laser, which form non-radiative recombination centres. The discussion of this paper drew attention to the comparable situation in GaP light-emitting diodes which has been overcome to some extent by surface passivation techniques.

The life problem was also mentioned by E. W. Williams (R.R.E., Malvern) when he suggested that InP may be a better material than GaAs as regards formation of defects which may influence degradation. However, it seems unlikely that an alloy system for use in possible heterostructure InP lasers is available with similar advantages to those of the $Al_xGa_{1-x}As$ system for the present GaAs lasers.

Several contributions to the conference dealt with the phenomena of filamentary action in semiconductor lasers at and just above threshold. The detailed experimental investigations of filaments reported by M. Mathews *et al.* (P.O.R.D., Dollis Hill) were analysed theoretically in contributions by G. H. B. Thompson (S.T.L., Harlow) and M. Cross (University College, Cardiff). The general concensus of opinion was that the filaments are a fundamental result of the shortened lifetime in the lasing region and *not* merely the result of inhomogeneities in the laser material. Although filamentary emission represents a problem for some applications, it may be turned to advantage for use in coupling to single-mode optical fibres in telecommunications systems. This was proposed by R. B. Dyott (P.O.R.D., Dollis Hill) who suggested applying an anti-reflective coating to the laser facet and simply placing the end of the fibre up against the facet. The refractive index of the fibre 'core' should be chosen to produce a suitable reflectivity at the laser mirror and hence encourage the formation of a lasing filament collinear with the 2 μm-diameter core. This would achieve an extremely efficient coupling (60% and more) of the laser output into the fibre, as compared with other techniques which are likely to only give a few per cent coupling efficiency. Other 'launching' possibilities dealt with by Dr. Dyott included distortion of the end of the fibre core to an elliptical shape roughly approximating to the rectangular facet of the laser whilst preserving the cross-sectional area.

The session on 'Applications' commenced with a review by M. Moncaster (S.R.D.E., Christchurch) on applications of GaAs lasers for defence purposes. This paper dealt mainly with multi-mode communications systems and was followed by two contributed papers on the same topic. Other contributions to this session included descriptions of high-power applications in range-finding and lidar systems.

The conference ended with a glimpse of the possible future of injection lasers in an invited paper on 'Future Prospects' by G. H. B. Thompson (S.T.L., Harlow). Dr. Thompson dealt with mode control in injection lasers: of transverse modes by the use of novel stripe-geometry and mesa lasers, and of longitudinal modes by means of external cavities. Mode-locking, by the use of external resonators and double-diode configurations, was also described, which can give trains of pulses with a present-day maximum repetition note corresponding to 500 ps between pulses. Dr. Thompson ended with proposals for an integrated optical system based on the $Al_xGa_{1-x}As$-GaAs epitaxial technology. The components envisaged included photodiodes, attenuators, superregenerative pulse amplifiers, modulators, and all-optical communications repeaters, all of which seem feasible as worthy additions to the family of the semiconductor injection laser.

M. J. ADAMS and B. THOMAS