Copy 2.

RESEARCH IN MICROWAVE ELECTRONICS

-by-

M. K. McPhun, B.Sc., M.I.E.E., C.Eng.

A thesis submitted for the degree of Doctor of Philosophy at the University of Warwick, England, May 1974.

Abstract

The thesis is a presentation of published work with a commentary. The 25 publications cover the period from 1961 to 1974, and with one exception (on the subject of silicon transistor circuits) are in the general area of microwave electronics.

Work on tunnel diode amplifiers includes a stability criterion, a measurement technique, amplifier designs for U.H.F. and X-band frequencies, and application to short hop radio-relay systems,

General purpose computer programs for the analysis of microwave circuits are discussed. The method of using chain matrices, which has since found wide application, is presented.

Work on microwave integrated circuits is presented which is mainly concerned with measurement techniques, and material and component properties. Measurement techniques are presented for thin dielectric films, rectangular dielectric plates, thin conducting metal films, and overlay capacitors. In each case microwave properties not previously obtained, are presented.

Declaration

The research discussed in sections 2.1., 2.2. and part of sections 2.3. and 2.4. was done prior to registration for the Ph.D. while I was employed by the U.K.A.E.A., the C.E.G.B., and Mullard Research Laboratories. This work was solely mine except where acknowledged. The contributions of respective authors to joint papers are made clear in the text. The remainder of sections 2.3. and 2.4. discuss research done subsequent to registration. This was carried out at the University of Warwick in the course of my work as a lecturer whilst supervising the research students who appear as co-authors of the papers submitted. In this work I have been an active participant, suggesting projects and obtaining the necessary facilities, engaging in regular discussions and contributing solutions to problems, and contributing to and editing the published papers. Further details of the contributions made by others is given in the Acknowledgements and in the main text.

Acknowledgements

At U.K.A.E.A., Capenhurst (now at Risley) I had many useful discussions with my group leader D.F. Davidson who introduced the work on silicon transistors, and with my colleagues E. Harrison, H. Jones, G. McGonigal, P. Roach, and A. Stead.

At C.E.R.L. Leatherhead my group leader J. Hooper inspired the work on tunnel diode repeaters. He and my colleagues S. Hartley, D.L. Hedderly, N. Prewitt, F. Pullen, and B.J. Terry all contributed to useful discussions.

At Mullard Research Laboratories, Salfords I was very grateful to my group leader C.S. Aitcheson for his perceptive and constructive criticism and the freedom he gave me, and to my colleagues J. Bandler, R. Davies, S.R. Longley, B.H. Newton, D.A.E. Roberts, K. Tweedale and J.F. Wells for many valuable discussions. At this time I also received useful advice from C.A.P. Foxall and his colleagues at A.S.M. Wembley (now part of G.E.C.). I am especially grateful to F.C. de Ronde of Philips Research Laboratories, Eindhoven for his advice on microwave measurement techniques and mechnical design, and for his personal interest and encouragement.

The work carried out at the University of Warwick would have been quite impossible without my research students R.S. Butlin, E.F. da Silva, B.G. Marchent, K. Mehmet, and D. Michie. They helped set up the elaborate facilities required for the work on microwave integrated circuits, and carried out most of the practical work of processing, measurement, writing computer

programs and drafts of joint papers. All my ideas have been subjected to their criticism, and much of the work and the solutions have arisen out of group discussions. J. Kelland, a final year undergraduate made a vital contribution at a critical time by commisioning our microphotography facilities. I am indebted to P.J. Smith for his painstaking work on the design and construction of a great variety of precision mechanical apparatus, and for contributing ideas on new construction techniques. H.V. Shurmer, S. Novak, J.M. Duckett, and E.G. Coates worked on setting up the Automatic Network Analyser and contributed to valuable discussions.

The research at Warwick was supported by S.R.C. grants B/SR/3536 and B/SR/9210, grants from the Clarkson Foundation and G.E.C., and a contract with R.R.E. Support for research students was given by the S.R.C., Microwave Associates and Redac Software Ltd. I am grateful to Professor A.L. Cullen who was a most helpful referee for the S.R.C. I have also had useful discussions with G.J. Hill of E.R.A., Leatherhead; Professor I.M. Stephenson and B. Easter of University College of North Wales, Bangor; J. Curran and J. Twistleton of G.E.C. Wembley, and R.F. Mayo of G.E.C. Stanmore (now at Lancaster University). Mullard Research Laboratories and the Plessey Company have been unstinting in their cooperation with helping me set up the processing facilities at Warwick.

In 1970 I spent 5 months sabatical leave at the Technishe Hogeschool Twente, Enschede, Netherlands. I wish to thank

Professor O.W. Memelink for providing the hospitality of his group and arranging for me to use the library and computing facilities.

In the course of the 14 years covered by this thesis there are many people not mentioned above who also have helped me. To them I am grateful and I am sorry that neither space nor my memory permit me to name them all.

I am most grateful to Professor J.A. Shercliff, chairman of the Department of Engineering, and Professor J.L. Douce for giving me such a free hand to develop microwaves as a major activity at Warwick, and making available a generous share of the available resources.

It is with regret that I record the death of Professor F.J. Hyde, my first supervisor, who encouraged me to take up this work at Warwick. I am grateful to Professor J.L. Douce, my present supervisor for negotiating the research grants with the Clarkson Foundation and G.E.C., and for advice on the preparation of this thesis.

Finally, the contribution of my wife Anne cannot be overstated as she has been consistent in her encouragement throughout, and in recent years frequently played the part of unpaid secretary and host.

CONTENTS

	Page
Declaration.	2
Acknowledgements.	3
Contents.	6
1. INTRODUCTION	9
2. COMMENTARY ON PUBLICATIONS	11
2.1. Direct Couping of Circuits using Silicon Transistors.	11
2.2. Tunnel Diode Amplifiers.	11
2.2.1. Future work.	14
2.3. Computer analysis of microwave circuits.	14
2.3.1. Future work.	15
2.4. Microwave Integrated Circuits.	16
2.4.1. Material measurements.	17
2.4.1.1. Future work.	18
2.4.2. Component measurements.	19
2.4.2.1. Future work.	20
3. CONCLUSIONS.	22
4. REFERENCES.	23

5. APPENDICES

A.1. Some advantages of silicon transistors in circuit design.

A.2. Operation of the tunnel diode above the resistive cutcff frequency.

A.3. Proposals for a 2-way tunnel diode repeater using a 4-port circulator.

A.4. Practical stability criterion for tunnel-diode circuits.

A.5. Practical stability criterion for tunnel-diode circuits.

A.6. Measurement of negetive resistance using a Z-g Diagraph.

A.7. U.H.F. tunnel-diode amplifier.

A.8. Short-hop radio-relay systems using tunnel-diode repeaters.

A.9. A tunable X-band tunnel diode amplifier.

A.10. Comparison of TEM with waveguide-below-cutoff resonators.

A.11. Microwave amplifier.

A.12. Variable-impedance transmission-line device.

A.13. Tunnel-diode amplifiers.

A.14. A computer programme for the analysis of branched distributed and lumped circuits.

A.15. Automatic general-purpose microwave circuit analysis programs.

A.16. A study of lumped components for microwave integrated circuits.

A.17. Microwave measurement of thin film dielectric properties.

A.18. A microwave relaxation phenomenon in electron-beam-evaporated films of SiO.

A.19. Thin film dielectric measurements.

A.20. Measurement of the properties of rectangular plates using a rectangular cavity.

A.21. Surface-resistance measurements of thin conducting films at 10 GHz.

A.22. Simple resonator method for measuring dispersion of microstrip.

A.23. Calibration of microwave network analyser for computer-corrected s parameter measurements.

A.24. Measurement of overlay capacitors at X-band with independent assessment of thin-film conductor losses.

A.25. The effect of parasitics on the design of overlay capacitors in microstrip circuits.

1. INTRODUCTION

This thesis is a presentation of published work, with a commentary. Each paper is presented as a separate appendix labelled and referenced A.1., A.2. etc. They are presented in logical, not chronologicsl order. Diagrams are referenced as Fig. A.2.5. for instance, meaning Fig. 5 of appendix A.2.

Chapter 2 is a commentary on the papers, which is divided into four sections. Sections 2.1., 2.2., and parts of sections 2.3. and 2.4. discuss research carried out at the U.K.A.E.A., C.E.R.L. and Mullard Research Laboratories. The remainder of section 2.3. and most of section 2.4. cover research carried out at the University of Warwick.

The following historical preview will show the connection between the papers. Section 2.1. on silicon transistors and circuits stands on its own. The research was done at the U.K.A.E.A. Central Instrumentation Laboratory while I was working on the control of industrial processes. All the other papers deal with aspects of microwaves. Section 2.2. deals with tunnel diode amplifiers. I set up this work in the U.H.F. band at C.E.R.L. and later moved to Mullard Research Laboratories to extend the work to X-band. The work on solid state amplifiers at X-band was found to be very dependent on the parasitics of the packages in which the active devices were mounted. Evidently there was a need to integrate the active device chips with the passive circuits, i.e. to make microwave integrated circuits (M.I.Cs).

I started this work at Mullard's and then moved to the University of Warwick and set up facilities for more fundamental work on lumped component M.I.Cs. This became my major research topic at Warwick. The research on the computer analysis of microwave circuits arose naturally out of the work on amplifiers. I wrote the first program at Mullard's because I needed to calculate frequency responses frequently. This program evidently filled a need of other workers, and at the University of Warwick I extended this work to form a research project in its own right.

2. COMMENTARY ON PUBLICATIONS.

2.1. Direct Coupling of Circuits Using Silicon Transistors.

The research described in appendix A.1. was done before silicon transistors were in common use. Silicon integrated circuits (S.I.Cs) were in the process of being invented, but methods of isolating collector regions of S.I.Cs. were not available.

The paper describes new circuit techniques available for use with silicon transistors, in particular the possibilities of direct coupoing. These techniques, e.g. Fig. A.1.8. have since become universally accepted and are normally used in S.I.Cs.

The basic amplifier circuit of Fig. A.1.4. is now used as a basic design example in undergraduate texts [1,2]. I presented this paper at three I.E.E. centres where I advocated the use of the amplifier circuit of Fig. A.1.4. in S.I.Cs. when the collector isolation problem had been overcome. This circuit was in fact later adopted for the Mullard S.I.C. type TAA 263.

2.2. Tunnel Diode Amplifiers.

Part of the work of the C.E.R.L. Telecommunications Group led by J. Hooper was directed towards establishing short hop microwave repeater links (A.8.). A suitable frequency band for operation of these links was 20 GHz to 30GHz, but solid state amplifiers in this frequency range did not exist at that time.

Tunnel diodes were just becoming available and could be used for amplifiers at the lower microwave frequencies. A number of workers were exploring ways of using the tunnel diode at much higher frequencies, and A.2. is part of that exploratory work. Hoffins and Ishii's reply is most unsatisfactory but I did not persue the discussion in public.

We had on hand a requirement for a repeater link at a frequency of 470 MHz. This was for the North-West of Scotland where a battery powered solid-state repeater was called for. Preliminary work at 470 MHz demonstrated the feasibility of using a tunnel diode repeater for this purpose, and I put forward the proposals of A.3. These proposals were accepted and funds provided for the development of a 2-way repeater at 470 MHz, and field trials.

The tunnel diode was the first wide-band negetive resistance device; it presents at its terminals an impedance with a real part which is negative over the frequency range from zero to the resistive cutoff frequency f_{ro}, and f_{ro} is typically a high microwave frequency. All workers in the field were hampered by undesired oscillations and lack of suitable measurement techniques. Suitable stability criteria did not exist for these essentially nonlinear negative resistance devices used in conjuction with distributed circuits. The stability criterion I evolved and subsequently used in this work is described in A.4. and A.5. The new measurement technique described in A.6. is equally suitable for use with the Hewlett Packard Network Analyser system at higher microwave frequencies.

The heart of the 470 MHz repeater consists of two identical reflection amplifiers which are described in A.7. These tunnel diode amplifiers were the first to incorporate independent control of diode bias point, gain, and centre frequency respectively. The whole repeater project is described in A.8. J. Hooper and I carried out the theoretical systems work, I designed the repeaters and planned the field trials initially, and D.L. Hedderly developed the repeaters and carried out the field trials, later taking over the whole project when I left C.E.R.L. in 1964. Subsequent work at C.E.R.L. on short-hop repeater systems has been successful, and such microwave links are now being assessed in full-scale field trials by the C.E.G.B. The main recommendations of A.8. have withstood the test of time and been adopted by the Post Office for their current work on short hop repeater systems[3].

My work at Mullard Research Laboratories was directed towards extending the upper frequency limit of tunnel diode amplifiers so as to produce useful amplifiers in X-band. The resulting amplifier uses similar design techniques to those of A.7. and is described in A.9. and A.11. It was exhibited at the Physics Exhibition, Alexander Palace, in 1966. This amplifier was uniquely simple to use; it was normal practise at that time for X-band tunnel diode amplifiers to have a multitude of tuning screws with an elaborate setting-up proceedure. The amplifier gain is controlled by a transmission line of variable characteristic impedance. This technique is further described in A.10. and A.12.

Design criteria for tunnel diode amplifiers, and the state-of-the-art in 1968 are summarized in A.13. Section 7.2. of A.13 was written by C.A.P. Foxall.

2.2.1. Future work.

The stability criterion of A.4. has still not been derived theoretically, and this would be worth doing. This criterion could probably be applied also to other negative resistance devices such as Gunn and avalanch diodes, which have superseded the tunnel diode as the main solid-state amplifying elements for microwave frequencies.

2.3. Computer Analysis of Microwave Circuits.

In the courseof the work on tunnel diode reflection amplifiers it was frequently necessary to compute frequency responses. Initially programs to do this for particular circuits (A.7.) were written to my specifications by the C.E.G.B. computing department. Later at Mullard Research Laboratories I wrote the general purpose program described in A.14. in which the circuit topology is defined by the input data. The method is based on 2-port chain matrices of networks which are cascaded or connected at tee junctions. This was the first work of this type. The program was further developed and used in the U.K. by Mullards, the Post Office, and the Plessey Company. In the U.S.A. the method was adopted by R.C.A. [4] and Sanders Associates [5].

Further philosophy of computing applied to microwave circuits is given in A.15. These desirable features of interactive programs using a graphical display were implemented at the University of Warwick by Marchent[6] under my supervision. He also carried out a detailed study of computation times which

shows that for microwave circuits the chain matrix technique is fastest even for highly interconnected circuits[6]. The disadvantage of this technique is that it cannot be used for circuits containing interactive loops, e.g. bridge types of interconnections cannot be solved. Marchent invented a new mixed matrix method of analysis capable of handling any network, with a reasonably short computation time [6,7]. He also rewrote the chain matrix analysis program in a form suitable for interactive computing. An important aspect of this work is that the computer sorts out the network topology, and decides the sequence of computation required [6]. Most of this work awaits publication.

2.3.1. Future work.

Any work on the computer optimisation of microwave circuits (A.15.) requires a suitable general purpose analysis program at its heart. High speed in this case is a prime requirement so the chain matrix method is very suitable. There remains the need to solve the problem of its use for circuits containing interactive loops. It is proposed that the chain matrix analysis be used to reduce the circuit to its simplest possible form. This would consist of single 2-port networks interconnected at junctions to form interactive loops. A typical microwave circuit would be most unlikely to have more than 10 junctions left. The solution would then be automatically continued to completion using Marchent's mixed matrix method. This would be fast as the number of junctions would be small. The concept is very simple, but implementing it presents difficult problems of topology.

Further work on incorporating a range of waveguide and microstrip circuit elements and discontinuities into these programs

would considerably increase their usefulness.

2.4. Microwave Integrated Circuits.

The performance of solid-state diodes in X-band is critically dependent on the parasitics of the diode packages. At Mullard Research Laboratoties we learnt to live with these parasitics and use them as circuit elements [8]. However the parasitics remained the limiting factor in determining the high frequency performance of many circuits. There was an evident need to integrate the semiconductor chips with the passive microwave components, i.e. to make hybrid microwave integrated circuits (M.I.Cs.).

The package parasitics are effectively minute lumped components even at microwave frequencies. Typical values are inductance in the range 0.1 to 1.0 nH and capacitance in the range 0.1 to 1.0 pF. These values are suitable for use as lumped components at microwave frequencies. I investigated the likely dimensions of such lumped components made by means of thin films deposited on insulating substrates, and concluded that they could be made easily using the photolithographic techniques developed within the semiconductor industry. I put forward proposals to Mullard's to start work on lumped component M.I.Cs. Before the proposals were accepted I was committed to move to the University of Warwick. However before leaving Mullard's, J.F. Wells joined me to make the first lumped component M.I.Cs. (A.16). The work at Mullard's continued after I left, concentrating mainly on applications [9]. I set up facilities at the University of Warwick with the aid of a grant from the S.R.C. and concentrated the work on material and component properties

for lumped component M.I.Cs.

The properties of thin films at microwave frequencies were quite unknown at that time, and no suitable measurement techniques existed. Neither were there measurement techniques available for the minute lumped components, and even microstrip measurements were confounded by transition discontinuities. Hence much of the work was involved with measurement techniques. Thorough surveys of the available materials, substrates, processes, and construction techniques for M.I.Cs. have been made by my research students [10,11,16,20].

2.4.1. Material measurements.

Considerable effort was devoted to developing a cavity technique for the measurement of thin dielectric films. The method is described in A.17 and A.19 and full details are given by Mehmet [10]. Noteworthy is the equivalent circuit method of calculating the loss of the films from the cavity return loss at resonance. This is simpler and more accurate than calculating the film loss from cavity Q measurements. This method was also used by Butlin [11] to optimise the r.f. sputtering process for depositing SiO_2 films. Some new results for thin films of SiO are given in A.18. A comprehensive review of the subject of thin film dielectric measurements is given in A.19.

In order to verify the cavity technique of A.17. for thin films we developed an independent method of measuring self-supported thin sheets of dielectric in a rectangular waveguide cavity [10]. We later applied this to the examination of the uniformity of M.I.C. substrates by measuring small samples cut

from the substrates. This work was done under contract to R.R.E. and later used by them [12]. The method is described in A.20. Subsequently I assisted E.R.A. to set up the same measurement technique, and it is now their standard method of assessing M.I.C. substrate materials [13].

A method for measuring conducting thin films was needed for assessing conductor losses in microstrip transmission lines and thin film capacitors. We evolved the cavity technique described in A.21., and further details are given by Butlin [11]. Again the technique uses an equivalent circuit approach yielding a better accuracy and simpler measurement proceedure than if Q measurements were used. We used the method to assess plated metal films [11], and some results obtained are given in A.24.

2.4.1.1. Future work.

We are now in a position to use our measurement methods to assess new techniques for depositing lower loss dielectric films. The present films are still too lossey. The r.f. sputtering technique Pitt [14] has used for low loss optical waveguides could well give the required results and should be investigated.

I hope to collaborate further with E.R.A. on dielectric measurements using the technique of A.20. An improved cavity has already shown promise. It is desirable to apply the equivalent circuit technique of A.17. to this work also. Further exploration of the applicability to these measurements of Lynch's relation [15] between dielectric constant and tanδ at r.f. and

microwave frequencies should be made. Discussions are also in progress with E.R.A. on the possibility of applying a microprocessor to the control of microwave dielectric measurements, and the computation of the results.

It should be possible to adapt the well-known perturbation formula of A.19. equation 2, into a form which enables the equivalent circuit technique of A.17. to be used with any cavity.

2.4.2. Component measurements.

X-band lumped components are typically less than 0.5 mm across, and we have spent much effort on developing special measurement techniques to overcome the errors caused by the transitions to normal measuring equipment. The most satisfactory so far is to mount the lumped component in a microstrip line for measurement purposes. We therefor needed to know accurate parameters for microstrip lines, and the new technique we used as a basis for obtaining these is described in A.22. Much further work on characterizing discontinuities in microstrip is described by Mehmet [10] and Michie [16].

The Hewlett Packard microwave network analyser system has featured prominently in our work on component measurement. The need to correct systematic errors both in the instrument and in the transitions to the unknown led me to propose an on-line system of computer correction. Simultaneously Hewlett Packard introduced their automatic network analyser (A.N.A.), but it was economical for us to develop our own using the existing computer within the Engineering Department. This also enabled us to

concentrate on the special facilities we required. G.E.C. seconded H.V. Shurmer to the university to work on the A.N.A. and support was also given by the Clarkson Foundation. Later I obtained further support for H.V. Shurmer from the S.R.C. who also supported a programmer. Initially the work was done using a G.E.C. computer [17,18,19]. The work has now been transfered to a Sigma 5 computer and is presently supported by another S.R.C. grant awarded to H.V. Shurmer. A major difficulty has been the construction of suitable calibration pieces, required in order to do measurements at a reference plane within microstrip. This has been overcome by obviating the need for a microstrip matched load, as described in A.23.

In order to measure overlay capacitors they were deposited directly into microstrip resonators [16]. Again the measurement method used an equivalent circuit approach and this is described in A.24. Critical to the use of any measurement technique for characterization of capacitors is a thorough assessment of the effect of the parasitics. This is described in A.25. which also presents design criteria for lumped overlay capacitors in X-band. Further details are given by Michie [16].

2.4.2.1. Further work.

We intend to combine the A.N.A. with the computer circuit analysis work of section 2.3. and static optimisation to produce a very powerful system for the C.A.D. of microwave circuits.

Much more work is required on reducing the loss of overlay

capacitors for use in X-band. This requires considerable technological support, and can only be undertaken if sufficient support is forthcoming.

Work is in progress on the construction and measurement of microwave lumped inductors [20]. This work is using a microstrip test jig and the measurement technique of A.23. This proceedure will also be suitable for the rapid characterization of a wide range of microstrip discontinuities.

3. CONCLUSIONS.

The published work is a record of research undertaken during the last 14 years, with one exception the papers being in the general area of microwave electronics. During the later half of this period 5 research students have been supervised by the candidate through work leading to the degree of Ph.D.

4) REFERENCES

1. D.L. SHILLING and C. BELOVE "Electronic circuits: discrete and integrated" McGraw-Hill, 1968, Ch 8.

2. P.E. GRAY andC.L.SEARLE "Electronic principles - physics, models, and circuits" John Wiley, 1969, Ch 16.

3. T.R. ROWBOTHAM "Short hop radio relay system work at 20 GHz" Proc. Microwave 73 Conference, Brighton, 1973, pp 112-116.

4. W.N. PARKER "DIPNET, a general distributed parameter network analysis program" I.E.E.E. Trans. MTT-17, 8, pp 495-505.

5. P.E. GREEN "General purpose programs for the frequency domain analysis of microwave circuits" ibid. pp 506-513.

6. B.G. MARCHENT "Interactive computer programs for the computer aided design of linear microwave circuits" PhD Thesis, University of Warwick, 1973.

7. B.G. MARCHENT "Computer aided design of microwave circuits" Proc. Summer School on Circuit Theory, Kobylisy, Czechoslovakia (Sept, 1971).

8. C.S. AITCHESON, R.DAVIES, and C.D. PAYNE "A balanced micropill-diode parametric amplifier (which supports the Q-band idling current in the series resonance of the varactor)" Proceedings of the 6th international conference on Microwave and Optical Generation and Amplification, Cambridge, September 1966, I.E.E. Conference Publication No. 27, pp 297-302.

9. C.S. AITCHESON et.al. "Lumped-circuit elements at microwave frequencies" I.E.E.E. Trans. MTT-19, 12, pp 928-937 (Dec. 1971).

10. K. MEHMET "Microwave integrated circuits - preparation and measurement techniques for materials and components" PhD Thesis, University of Warwick, 1970.

11. R.S. BUTLIN "Microwave integrated circuits- preparation and measurement techniques for metal and dielectric films" PhD Thesis, University of Warwick, 1973.

12. J.R. BOSNELL "The effect of the physical properties of alumina substrates on their use in microwave hybrid circuits" Microelectronics, 3, 10, pp 33-39, (1971).

13. G.J. HILL "Alumina for microwave applications" in High Frequency Dielectric Measurement, I.P.C. Science and Technology Press Ltd, pp 127-134, (1973).

14. C.W. PITT "Sputtered-glass optical waveguides" Electronics Letters, 9, 17, pp 410-403 (23rd Aug. 1973).

15. A.C. LYNCH "Relationship between permittivity and loss tangent" Proc. I.E.E., 118, 1, pp244-246, (Jan. 1971).

16. D. MICHIE "Microwave integrated circuits - preparation and measurement techniques for overlay capacitors" PhD Thesis, University of Warwick, 1974.

17. H.V. SHURMER "Correction of a Smith chart display through bilinear transformations" Electronics Letters, 5, 10, p209, (15th May 1969).

18. H.V. SHURMER "New method of calibrating a network analyser" Electronics Letters, 6, 23, pp 733-734, (12th Nov. 1970).

19. H.V. SHURMER "Low-level programming for the on-line correction of microwave measurements" Radio and Electronic Engineer 41, 8, pp 357-364, (Aug. 1971).

20. E.F. da SILVA "Inductors as lumped elements in microwave integrated circuits" MSc Thesis, University of Warwick, 1971.

APPENDIX 1.

Some advantages of silicon transistors in circuit design.

Prodeedings I.E.E. 108B, 41, pp570-575, (Sept 1961).

THE INSTITUTION OF
ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

SAVOY PLACE, LONDON, W.C.2

SOME ADVANTAGES OF SILICON TRANSISTORS IN CIRCUIT DESIGN

By

M. K. McPHUN, B.Sc.(Eng.), Graduate

Reprint from

THE PROCEEDINGS OF THE INSTITUTION, VOL. 108, PART B, NO. 41, SEPTEMBER 1961

SOME ADVANTAGES OF SILICON TRANSISTORS IN CIRCUIT DESIGN

621.382.3

The Institution of Electrical Engineers
Paper No. 3671 M
Sept. 1961

SOME ADVANTAGES OF SILICON TRANSISTORS IN CIRCUIT DESIGN

By M. K. McPHUN, B.Sc.(Eng.), Graduate.

(*The paper was first received* 10*th March, and in revised form* 17*th May*, 1961.)

SUMMARY

After transistors had been used for industrial instrumentation within the Development and Engineering Group of the United Kingdom Atomic Energy Authority for some years, silicon transistors were adopted for general-purpose use, as they were expected to be more reliable than germanium transistors. It has since become apparent that they possess more advantages over germanium transistors than is generally recognized, or was thought at the time of their adoption. For example, direct coupling of transistor circuits is facilitated and the number of components required may be much reduced; the use of electrolytic capacitors may be avoided.

The paper compares the performance of silicon and germanium transistors from the viewpoint of the circuit designer; physical origins of their characteristics are not discussed. A range of circuits which take advantage of the characteristics peculiar to silicon transistors is described. The circuits are for direct-coupled amplifiers, current amplifiers for small signals, switching applications and multivibrators with long periods.

(1) INTRODUCTION

Since the advent of silicon transistors, descriptions have been given of their use for special purposes, e.g. choppers[1] and low-drift input stages for d.c. amplifiers.[2]

Reliability is of prime importance in industrial control equipment, particularly in the atomic energy industry. As soon as an assured supply of silicon transistors became available, it was decided to use them instead of germanium transistors for general purposes in the Central Instrument Laboratory of the U.K.A.E.A. (Development and Engineering Group) for the following reasons:

(*a*) They were expected to be more reliable than germanium transistors.
(*b*) They could be operated at higher temperatures than germanium transistors, a particularly useful feature in chemical plants.
(*c*) The manufacturers predicted that their costs would fall below those of germanium transistors within a few years, making germanium transistors obsolescent for general purposes.

The paper deals with experience gained in using silicon transistors for a wide variety of applications. The design of the individual circuits for these applications is not described in full where the techniques used are conventional; only those points of special relevance to the use of silicon transistors are described.

(2) COMPARATIVE PERFORMANCE OR GERMANIUM AND AVAILABLE SILICON TRANSISTORS

This discussion is concerned with transistor performance as it affects the circuit designer; the physical origins of the characteristics are not considered.

(2.1) Input Characteristics

The current/voltage curves for the emitter-base diode of typical germanium and silicon alloy-junction transistors (Fig. 1) show a sharper turnover for the silicon than for the germanium transistor. As with the forward-biased silicon diode, the forward voltage across the emitter-base diode may be considered as constant at 0·6 V. For applied voltages of less than 0·4 V the junction may be considered to be reverse-biased. This forward bias necessary for conduction may not be ignored in circuit design as often as it is with germanium transistors; it may, however, be put to good use, as is shown throughout the paper.

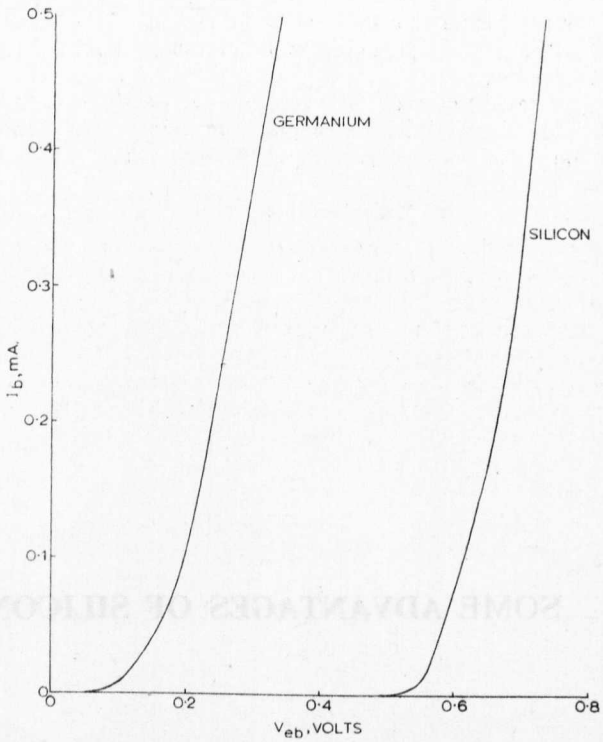


Fig. 1.—Typical characteristics of emitter-base diodes for alloy-junction germanium and silicon transistors.

(2.2) Leakage Current and Temperature

The collector current flowing with open-circuit emitter is much smaller for the silicon than for the germanium transistor. This is the common-base leakage current, I_{co}; the published applications[1, 2] of silicon transistors make use of the fact that I_{co} is very small so that its dependence on temperature has much less effect on circuit performance than with the germanium transistor. The effect of temperature on the leakage currents is the same for both germanium and silicon transistors, i.e. an increase in temperature of about 9° C causes the leakage current to double; Table 1 shows typical figures for silicon and germanium alloy-junction transistors.

In many general-purpose circuits the leakage current of the silicon transistor may be completely neglected, thus obviating the need for stabilizing the d.c. working point against changes in leakage current with temperature; this leads to simpler circuits and greater reliability.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. McPhun was formerly with the United Kingdom Atomic Energy Authority (Development and Engineering Group) and is now at the Central Electricity Research Laboratories.

Table 1

COMPARISON OF LEAKAGE CURRENTS AT 25° C FOR GERMANIUM AND SILICON TRANSISTORS

Description	Type	I_{co}	I'_{co}
		μA	μA
A.F. germanium	OC71	5	150
R.F. germanium	OC44	0·5	25
Silicon	OC201	10^{-3}	60×10^{-3}

The other main temperature effect is the variation of V_c with temperature at constant collector current. This effect is the same for both silicon and germanium transistors, being about 2·5 mV per deg C.

The permissible range of operating temperatures is much greater for silicon than for germanium transistors. Typical maximum operating temperatures are germanium, 75° C and silicon, 150° C.

(2.3) Collector Resistance

Measured values of common-base collector resistance, r_c, at normal operating voltages of the order of 6 V are lower for typical alloy-junction silicon transistors than for similar germanium types, but they are markedly higher at low values of collector-base voltage, V_{cb}.

The characteristic of the collector-base diode of the simple equivalent circuit shown in Fig. 2 is similar to that of the

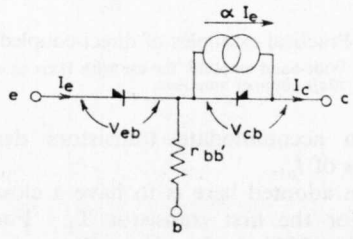


Fig. 2.—Simple equivalent circuit of transistor.

emitter-base diode, Fig. 1. For normal operation as an amplifier, the emitter-base diode is forward-biased and the collector-base diode is reverse-biased. However, the silicon diode is effectively reverse-biased for voltages through zero to forward values of a few tenths of a volt, and it would thus be expected that transistor action would be maintained with high collector resistances for collector-base voltages in this range. This is indeed so, as shown in Fig. 3 which gives measured curves of r_c plotted against V_{cb} for both germanium and silicon transistors. Over the range 0·2 V reverse to 0·1 V forward for V_{cb}, the collector resistance of the silicon transistor decreases by one-third while that of the germanium transistor decreases by three decades.

(2.4) Noise

Measurements[3] showed that the best type OC200 silicon transistors were slightly less noisy than the best available germanium transistors. The manufacturers claim the silicon OC201 to have a lower noise figure than the OC200, and measurements show a noise figure of 6 dB with a source resistance of 500 Ω to be easily obtainable at 1 kc/s with the OC201.

The OC201 was adopted for general-purpose use, and experience has shown that its noise figure is comparable with that of the best available germanium transistors.

A lower noise figure is obtained by operating a transistor at very low values of collector current.[3] This is facilitated with silicon transistors, as collector currents of 50 μA or less may be used in low-noise circuits without I'_{co} proving troublesome; under these conditions linearity may be a limiting factor.

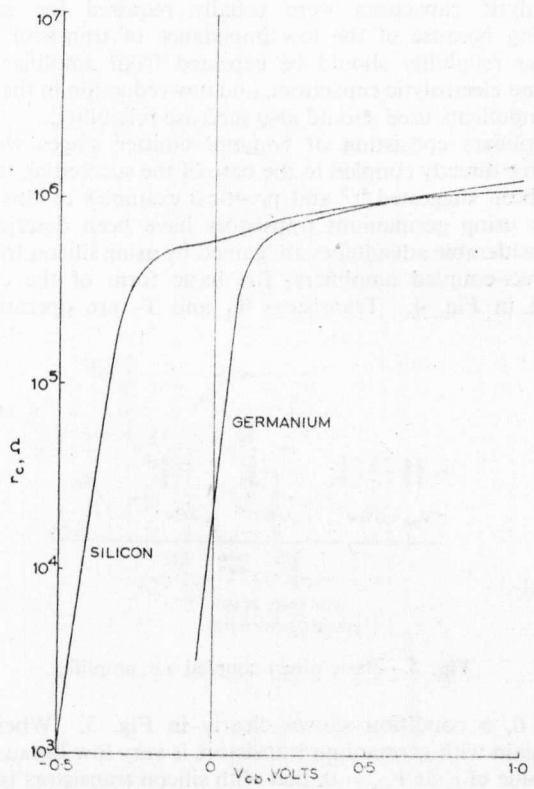


Fig. 3.—Collector resistance as a function of collector-base voltage for typical alloy-junction germanium and silicon transistors.
Collector current = 1 mA.

(2.5) Availability of Silicon Transistors and Quantitative Data

At present there are few types of silicon transistors available in this country;* all dissipate about 200 mW at room temperature and have common-base cut-off frequencies between 1 and 4 Mc/s. The germanium transistors of similar size and performance dissipate only 100 mW at room temperature because of their lower maximum junction temperature.

Silicon power and high-frequency transistors are now becoming available, but their price restricts their use to special applications. The cut-off frequency of the alloy-junction silicon transistors, however, has been found adequate for most instrument work, the exception being due to the difficulty in making stable wide-band amplifiers with large amounts of overall negative feedback.

So far a very large spread in the characteristics of silicon transistors must be tolerated; values of the common-emitter current gain α' from 20 to 130 may be expected in a batch of 20 transistors. New types with a narrow spread of α' have recently become available, but even with these, values of α' may still be found in the range 30–75.

Quantitative information about silicon transistors is scarce, and this puts circuit designers at a disadvantage. Most characteristics must be measured by the user, and because of the large spread this can be a very laborious process. For this reason, most of the circuits described here will tolerate a very wide spread of transistor characteristics.

(3) DIRECT-COUPLED AMPLIFIERS

The frequent requirement for 50 c/s servo amplifiers showed the need for a.c. amplifiers using direct coupling between stages;

* A rapid increase in the variety of silicon transistors available has taken place since the paper was written.

electrolytic capacitors were usually required for interstage coupling because of the low impedance of transistor circuits. Greater reliability should be expected from amplifier circuits using no electrolytic capacitors, and any reduction in the number of components used should also increase reliability.

Amplifiers consisting of common-emitter stages with each collector directly coupled to the base of the succeeding transistor have been suggested,[4, 5] and practical examples of this type of circuit using germanium transistors have been described.[1, 4, 5]

Considerable advantages are gained by using silicon transistors in direct-coupled amplifiers; the basic form of the circuit is shown in Fig. 4. Transistors T_1 and T_2 are operating with

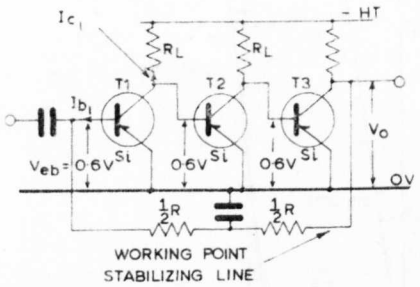


Fig. 4.—Basic direct-coupled a.c. amplifier.

$V_{cb} = 0$, a condition shown clearly in Fig. 3. Whereas the stage gain with germanium transistors is very low because of the low value of r_c at $V_{cb} = 0$, that with silicon transistors is normal for the case where $r_c(1 - \alpha) > R_L$.

Let us consider germanium transistors used in this circuit.[5] The collector resistance is reduced to one-hundredth of the value obtained with a few volts applied to the collector, then $r_c(1 - \alpha) \ll R_L$, and this severely limits the gain per stage; moreover, under these conditions r_c is very sensitive to temperature, and so also is the stage gain. The voltage swing at a collector must be limited to the millivolt region because of the change in r_c with V_{cb}. A voltage-divider coupling must be used to the last stage[4] in order to allow sufficient voltage swing at the base of T_3; this, in turn, means that a positive supply rail must be used.

With silicon transistors, a high value of r_c is maintained, and so $R_L \ll r_c(1 - \alpha)$; the stage gain is not then limited by r_c and the small variations in r_c with V_{cb} have negligible effect. The maximum voltage swing at a collector may be of the order of 200 mV; this is sufficient to provide a signal at the T_3 collector with a peak voltage of a few volts, no positive rail being required for an interstage coupling. Further, as the stage gain is governed by R_L and not r_c, any temperature-dependence of r_c does not affect the gain.

Table 2

EFFECT OF GAIN OF T_1 ON V_o (FIG. 4)

α_1'	I_{b1}	V_o
	mA	V
30	1/30	10
40	1/40	7·5
60	1/60	5

The simple operating-point stabilization shown in Fig. 4 is adequate for most purposes. The required base direct current, I_{b1}, of the first stage is determined; R is then chosen so that $V_o - V_{eb} = I_{b1}R$. Variations of I'_{co} with temperature may be neglected as this current is so small; variations of V_{eb} with temperature are much less than V_o. Difficulty is encountered, however, if it is desired to have a close control on V_o, and at the same time to accommodate transistors demanding widely differing values of I_{b1}.

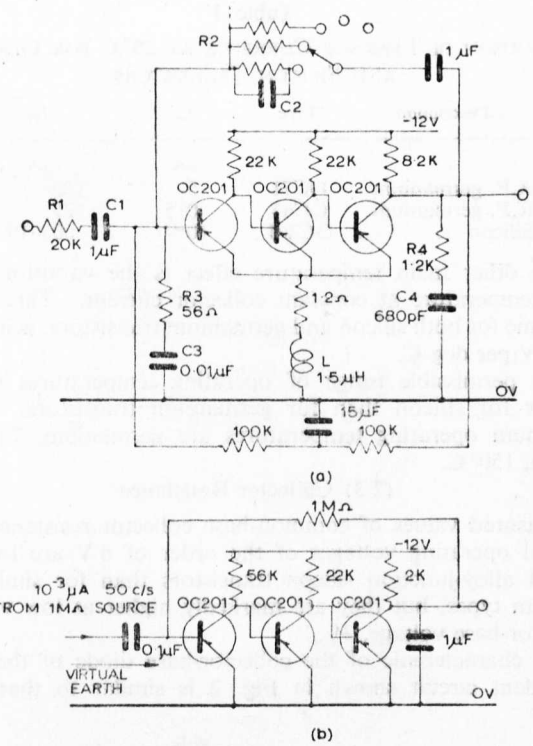


Fig. 5.—Practical examples of direct-coupled amplifiers.
(*a*) Wide-band amplifier for use with 1 kc/s square waves.
(*b*) 50 c/s chopper amplifier.

The solution adopted here is to have a close control on the value of α_1' for the first transistor T_1. For example, with $I_{c1} = 1$ mA, $R = 300$ kΩ, the values shown in Table 2 will be obtained. The variations of α_1' tolerated will be governed by the permissible variation of V_o.

Practical examples of these circuits are shown in Fig. 5. Fig. 5(*a*) shows a wide-band amplifier for use with 1 kc/s square waves. Fixed current feedback is applied, and gains of 1, 2, 5, 10, 20 and 25 are obtained by switching the shunt feedback resistors. Stabilizing networks R_3C_3, R_4C_4, C_2 and L are necessary because of the large degree of feedback and wide bandwidth of 100 kc/s.

For the low-noise chopper amplifier shown in Fig. 5(*b*) it was convenient to use the same resistor for feedback at the 50 c/s signal frequency as for stabilizing the d.c. working point. An electrolytic capacitor would have been required for decoupling a separate working-point stabilization loop at 50 c/s.

(4) CURRENT AMPLIFIERS FOR SMALL SIGNALS

It is sometimes inconvenient to use an electrometer valve in conjunction with equipment otherwise using transistors, usually because the power supplies required are not available. Current amplifiers using silicon transistors have been successfully used in such applications and two examples are described.

Fig. 6 shows the circuit of a head amplifier for a photomultiplier detecting very low levels of light chopped at 1 kc/s. Two cascaded emitter followers are used, giving a current gain of about 300. The collector current of T_1 is set at 120 µA. This is a compromise between the low value required for low noise and the need for linearity over a range of input currents of three decades, from 10^{-8} to 10^{-11} A peak. The noise generated in the first transistor is negligible compared with the shot noise of

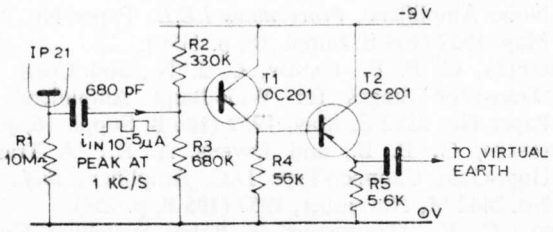


Fig. 6.—Circuit of photomultiplier head amplifier.
The type IP21 photomultiplier is selected for low noise.

the photomultiplier. It is possible to observe individual electrons leaving the photomultiplier cathode; this suggests that the circuit could be used with counting equipment for photon counting at very low light levels.

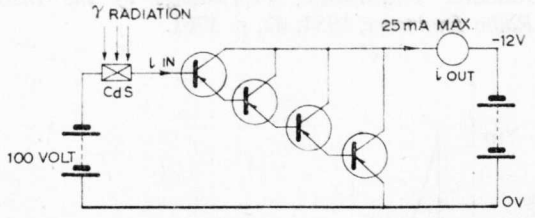


Fig. 7.—Circuit of amplifier for crystal-type radiation detector.
Transistors are type OC201.

Fig. 7 shows the circuit of an amplifier for direct currents suitable for use with crystal radiation detectors. A typical cadmium-sulphide detector in this circuit will pass current, i_{in}, up to 2 μA. The minimum current gain is $1{\cdot}3 \times 10^7$. The total leakage current at the output is 1·3 μA at 25° C; thus, at room temperature, operation over a range of four decades is possible for values of i_{out} from 2·5 μA to 25 mA. The linearity over this range is only 10% but this is unimportant in this application as the crystal radiation detectors are non-linear.

(5) SWITCHING APPLICATIONS

Since a forward bias of 400 mV is required at the base of a silicon transistor to start conduction, it is easy to ensure that it does not conduct; it is never necessary to make the base positive with respect to the emitter to reduce the collector current to below 1 μA at room temperature, and thus no positive supply rail is required.

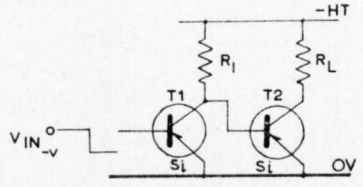


Fig. 8.—Direct-coupled switching circuit.

In the circuit of Fig. 8, T_2 may be adequately cut off by bottoming the collector of T_1; there is a clear margin between the bottoming voltage of T_1, typically 100 mV, and the 450 mV required at the base of T_2 for conduction.

(5.1) Relay-Drive Circuits

A relay may be substituted for R_L in Fig. 8, if protection is provided for T_2 against voltage overshoot when T_2 is being cut off.[6] Apart from ensuring that transistor ratings are not exceeded, the only design requirement is that the current through R_1 with T_1 cut off is sufficient to bottom T_2.

If sufficient voltage swing is available to drive the relay and current amplification only is required, the circuit of Fig. 9 may

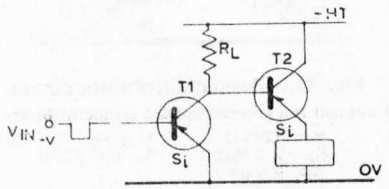


Fig. 9.—Relay-drive circuit.

be used. With T_1 cut off, T_2 is saturated. When T_1 is bottomed, T_2 will continue to conduct at first as the stored energy of the relay when released causes conduction of the emitter-base junction. Thus, no surge protection is required for T_2 but the relay is necessarily slow to release.

When it is desirable to operate T_1 with the minimum possible collector current, a more efficient circuit is obtained by using all the collector current of T_1 as base current for T_2. The circuit of Fig. 10 does this, using an *n–p–n* transistor for T_2.

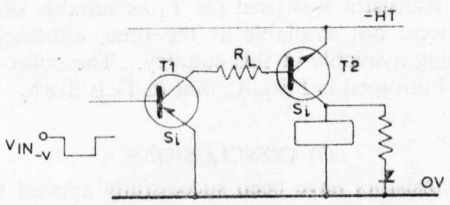


Fig. 10.—Efficient relay-drive circuit.

Both T_1 and T_2 are cut off simultaneously and bottomed simultaneously. The value of R_L fixes the T_1 collector current when T_1 is bottomed, this being made equal to the base current required to bottom T_2.

(6) MULTIVIBRATOR WITH LONG PERIOD

It is now generally recognized that a better astable multivibrator is obtained by placing the timing circuits in the emitters instead of in the collectors of transistors[7, 8]; when both junctions are reverse-biased the emitter leakage current is much less than the collector leakage current and I_{co} is the sum of these two.[9] However, there is an exception when long periodic times are required, since the emitter-timed circuit requires larger timing capacitances than does the base-timed circuit. To avoid the use of electrolytic capacitors for the timing circuits it is necessary to keep the timing capacitance as small as possible.

The small leakage current of silicon transistors allows the use of the base timing circuit, and also high impedances. It is shown in Section 10 that the maximum usable base timing resistance is obtained by returning it to an infinite aiming potential, V (Fig. 11). Little is gained by increasing V beyond $2V_{HT}$, and unless the additional supply rail is available, V will be made equal to V_{HT}. Fig. 11 also gives the component values required for a symmetrical astable multivibrator with a periodic time of 2 sec. The collector current when bottomed is only 100 μA, requiring a base current of 5 μA.

An asymmetrical circuit used to drive a tape reader via a relay at a frequency of 0·5 c/s is shown in Fig. 12. An asymmetrical circuit was needed to provide sufficient base current to the relay-drive transistor, T_3. The relay-drive circuit of Fig. 10

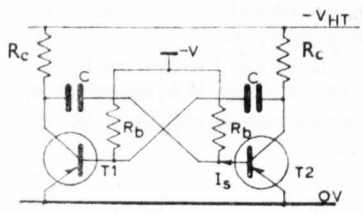


Fig. 11.—Basic multivibrator circuit.
The circuit is a symmetrical 0·5 c/s multivibrator with

R_c: 120 kΩ	$V = V_{HT} = -12$ V
R_b: 2·2 MΩ	T_1, T_2: OC202
C: 0·6 μF	

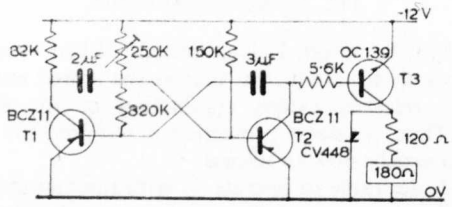


Fig. 12.—Asymmetrical multivibrator for driving a tape reader.

was used to give the greatest economy of current in T_2. A germanium transistor was used for T_3 as suitable silicon *n–p–n* transistors were not available at the time, although they are now becoming available in this country. The collector current in T_1 when bottomed is 140 μA; that in T_2 is 2 mA.

(7) CONCLUSIONS

Silicon transistors have been successfully applied to general-purpose use in industrial instrumentation to give increased reliability; none has yet failed in use in the experience of the laboratory.

The use of silicon transistors has led to simpler circuits with fewer components than if germanium transistors had been used, and the process of circuit design has been simplified. In particular, the use of electrolytic capacitors has been avoided, leading again to increased reliability.

The predicted decrease in cost of silicon transistors has occurred, to the extent of about one-half.

(8) ACKNOWLEDGMENTS

Mr. D. F. Davidson, of the Central Instrumentation Laboratory, Capenhurst, was responsible for introducing silicon transistors for general-purpose use in the laboratory and suggested their use in direct-coupled circuits. Helpful discussions were held with Mr. G. McGonigal of the same laboratory, and Mr. H. Jones developed the circuit shown in Fig. 7.

The paper is published by permission of Sir William Cook, Managing Director, and Dr. H. Kronberger, Deputy Managing Director, of the United Kingdom Atomic Energy Authority (Development and Engineering Group).

(9) REFERENCES

(1) Hutcheon, I. C., and Summers, D.: 'A Low-Drift Transistor Chopper-Type D.C. Amplifier with High Gain and Large Dynamic Range', *Proceedings I.E.E.*, Paper No. 3227 M, March, 1960 (**107** B, p. 451).
(See also Nambiar, K. P. P.: Discussion on the above paper, p. 461.)

(2) Kemhadjian, H.: 'Transistor Amplifiers for D.C. Signals', *Mullard Technical Communications*, 1958, **4**, p. 162.

(3) Battye, C. K., and George, R. E.: 'Transistors as Low-Noise Amplifiers', *Proceedings I.E.E.*, Paper No. 2933 E, May, 1959 (**106** B, Suppl. 18, p. 1190).

(4) Chaplin, G. B. B., Candy, C. J. N., and Cole, A. J.: 'Transistor Stages for Wide-Band Amplifiers', *ibid.*, Paper No. 2892 E, May, 1959 (**106** B, Suppl. 16, p. 762).

(5) Chaplin, G. B. B., and Owens, A. R.: 'A Transistor High-Gain Chopper-Type D.C. Amplifier', *ibid.*, Paper No. 2442 M, November, 1957 (**105** B, p. 258).

(6) Hill, C. F.: 'Transistors in Relay Switching Circuits', *Mullard Technical Communications*, 1956, **2**, p. 284.

(7) Bowes, R. C.: 'Timing Circuits and Waveform Generators', *R.R.E. Journal*, April, 1960, No. 44, p. 121.

(8) Bowes, R. C.: 'A New Linear Delay Circuit based on an Emitter-Coupled Multivibrator', *Proceedings I.E.E.*, Paper No. 2951 E, May, 1959 (**106** B, Suppl. 16, p. 793).

(9) Ebers, J. J., and Moll, J. L.: 'Large-Signal Behaviour of Junction Transistors', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1761.

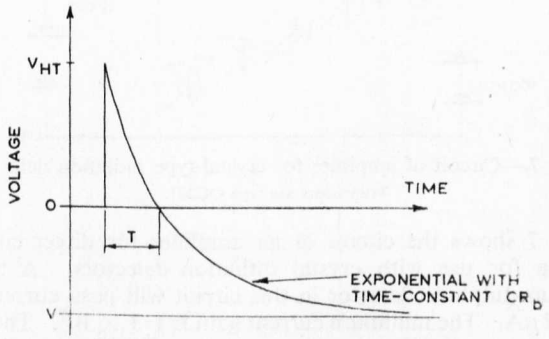


Fig. 13.—Timing action of multivibrator of Fig. 11.

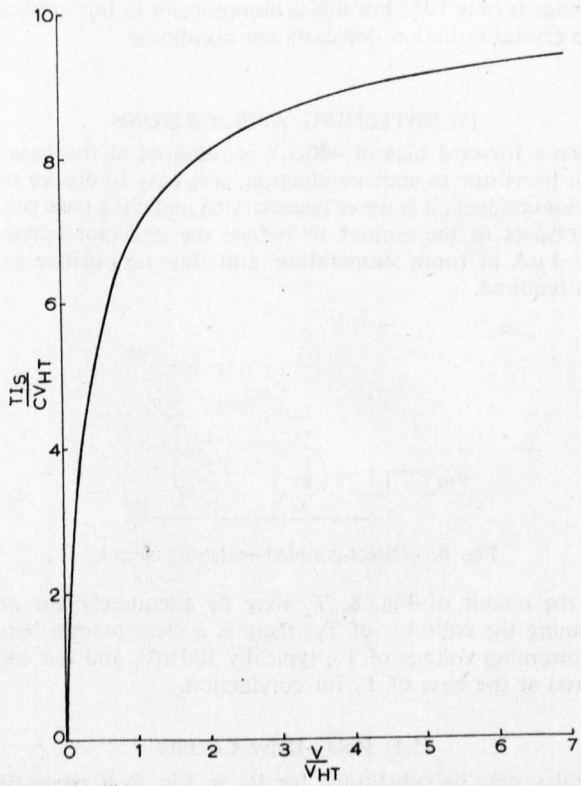


Fig. 14.—Calculated curve of TI_S/CV_{HT} as a function of V/V_{HT}.

(10) APPENDIX: MAXIMUM TIMING RESISTANCE FOR A MULTIVIBRATOR WITH LONG PERIOD

Referring to the circuit of Fig. 11, we require the value of V to give the maximum value of R_b. The current I_S required to bottom the transistor will be determined by R_b and V, which both enter also into the expression for the timing operation.

Let the time that a transistor is cut off be T. The timing action is shown in Fig. 13. The bottoming voltage is neglected and the transistor is assumed to conduct as soon as its base goes negative.

It is known that $T = CR_b \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$

We have also $R_b = \frac{V}{I_S}$

Eliminating R_b, we obtain

$$T = \frac{CV}{I_S} \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$$

In terms of V_{HT}/V as a variable,

$$\frac{TI_S}{CV_{HT}} = \frac{V}{V_{HT}} \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$$

In Fig. 14 is shown a dimensionless plot of TI_S/CV_{HT} against the ratio V/V_{HT}. Little is gained by increasing V beyond $2V_{HT}$.

Printed in Great Britain by
UNWIN BROTHERS LIMITED, WOKING AND LONDON

APPENDIX 2.

Operation of the tunnel diode above the resistive cutoff frequency.

Proceedings I.E.E.E. 51, 9, pp 1265-1267 (Sept. 1963)

Operation of the Tunnel Diode Above the Resistive Cutoff Frequency*

The waveguide mount described by Hoffins and Ishii[1] enabled them to operate a tunnel diode at frequencies about three times the resistive cutoff frequency by "absorbing the junction capacity and series inductance of the diode into the distributed parameters of the waveguide," but the authors do not attempt to explain how this "absorption" occurs. The following is offered as a tentative explanation of this phenomenon and examines its practical implications. Discussion would be welcome.

The tunnel diode equivalent circuit of Fig. 1 is well established,[2,3] as is the resistive cutoff frequency

$$f_0 = \frac{1}{2\pi CR}\sqrt{\frac{R}{r} - 1}$$

above which the impedance looking into the tunnel diode terminals 1 and 2 is positive. No passive circuit connected to terminals 1, 2 can result in oscillations at frequencies in excess of f_0; hence, if power at such frequencies is to be obtained, one of the following explanations must apply.

1) The power observed was a harmonic of an oscillation on a fundamental frequency less than f_0.
2) The waveguide mount did not present

* Received May 1, 1963.

[1] C. C. Hoffins and K. Ishii, "Microwave tunnel-diode operation beyond cut-off frequency," PROC. IRE (*Correspondence*), vol. 51, pp. 370–371; February, 1963.

[2] R. A. Pucel, "Physical principles of the Esaki diode and some of its properties as a circuit element," *Solid State Electronics*, vol. 1, pp. 22–23; March, 1960.

[3] H. Fukui, "Microwave figure of merit for tunnel diodes," *Electronics*, vol. 34, pp. 62–64; March 3, 1961.

a passive impedance to the tunnel diode.

3) The tunnel diode equivalent circuit is not valid at the frequencies concerned.

4) Access has been obtained to within the equivalent circuit.

Explanation 2) is most unlikely, so presuming the authors took the obvious step of ensuring 1) was not taking place, we are left to examine explanations 3) and 4). The equivalent circuit has been examined in detail,[2,3] and the evidence is that it is valid to well beyond f_0; refinements such as considering skin effect certainly give no cause for believing that there are frequencies above f_0 for which a negative resistance could be obtained between terminals 1, 2. We are left with explanation 4).

If we had access to terminals 3, 4 there would be no inherent upper limit to the frequency of oscillation apart from that imposed by the speed of tunnelling within the junction—frequencies of order 50,000 Gc! To operate at frequencies above f_0 then, we must go at least some way towards connecting the external circuit to terminals 3, 4, *i.e.*, by "getting inside" the series resistance r. It would seem that this waveguide mount "gets inside" r by allowing the electric field within the waveguide to interact with the tunnel junction directly, *i.e.*, the concept of the waveguide presenting "an impedance" to the recognized terminals on the case of the tunnel diode (terminals 1, 2 of Fig. 1) must be abandoned.

If this concept is abandoned, a new light is thrown upon both the design of waveguide mounts, and upon the tunnel diode mounts themselves.

We would expect to obtain the maximum benefit from the tunnel diode when the greatest part of the electric field in the waveguide is concentrated at the tunnel junction. This suggests the use of a very narrow dielectric-filled waveguide of much reduced height, interrupted by the diode.

Now consider the diodes. The present technique of constructing microwave tunnel diodes accepts the presence of r as inevitable and concentrates on reducing C by etching away the junction area.[4] This is worthwhile and increases f_0 in spite of an increase in r. The resulting diodes are very low power devices, and have a shape similar to Fig. 2. Due to the pinch effect on the current, most of the series resistance r occurs in a region very close to the junction. A junction of this shape does not seem very promising where "getting inside" r is concerned. If, however, we do not consider r as an inevitable barrier to high frequency operation, there then is no point in etching away the junction to reduce C; r is no longer concentrated in the region of the junction, hence, the possibility of "getting inside" r is increased. Such a diode in a suitable waveguide mount may have power handling capabilities far in excess of the presently accepted limits for the higher microwave frequencies.

Summarizing, the author believes that if it is possible to operate a tunnel diode in a waveguide mount above the resistive cutoff frequency, then the concept of impedance must be abandoned when considering such circuits. If this is done, then the design of tunnel diodes for operation at high microwave frequencies may have to be radically reviewed.

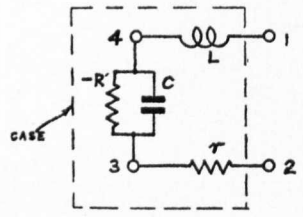


Fig. 1—Tunnel diode equivalent circuit.

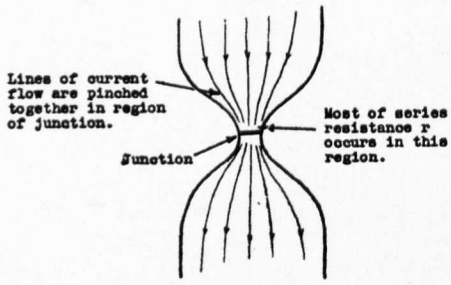


Fig. 2—Shape of junction.

M. K. McPhun
Central Electricity
Research Labs.
Cleeve Road
Leatherhead, Surrey,
England

Authors' Comment[5]

1) The authors agree with McPhun's remarks. This is essentially what the authors are trying to show by describing the authors' waveguide mounts.[6,7,8] A diode with a low peak current was used due to its availability and ease of biasing. The authors feel that a higher power diode with a larger junction capacity would perform equally well even though this was not investigated.

2) The statement, "If, however, we do not consider r as an inevitable barrier to high frequency operation there then is no point in etching away the junction to reduce C," may be true, but the authors are not quite sure about the explanation applying. This would require considerable investigation with various diodes and a theoretical analysis.

3) It is true that the equivalent circuit shown in Fig. 1 of McPhun's remarks is well established. No matter what frequency range is concerned, the equivalent circuit configuration would not be changed. It should be noted, however, that the equivalent quantities of L, C, r and $-R$ are frequency dependent and the values determined by a low-frequency method are not likely to be applicable at microwave frequencies.[9]

4) An exact solution can be obtained by solving wave equations with the complicated boundary values of actual microwave tunnel diode mount. The method is difficult, complicated and very involved. Therefore, the authors will not do it this time. Instead of showing the exact solution, the authors will explain how "absorbing the junction capacity and series inductance of the diode into the distributed parameters of waveguide"[8] is possible. An idealized model will be used instead of the actual tunnel diode mount to merely *explain* the possibility of "absorption."

5) In microwaves, it is not difficult to find the "access" to terminals 3 and 4 in McPhun's Fig. 1. McPhun's Fig. 1 is transplanted in the authors' Fig. 3. In the authors' Fig. 3, inside the "case" is the tunnel diode. If, at terminals 1 and 2, a transmission line of distribution parameter R, C, r, and L is connected, L, C and r within the tunnel diode case becomes a part of the transmission line and thus "the junction capacitance and series inductance of the diode" is "absorbed into the distributed parameter of waveguide." If the transmission line in Fig. 3 represents a waveguide, r represents the ohmic resistance per unit length of the line corresponding to the longitudinal conduction current path, R represents the resistance corresponding to the transverse displacement and conduction current paths, C represents the capacitance per unit length of the line along the transverse displacement current path and L represents the inductance per unit length of the line along the longitudinal conduction current path.[10] It was found that the order of r in the waveguide is almost of the same order as r in the tunnel diode, though the value of r in the tunnel diode at high frequencies becomes a distributed parameter having a configuration which is not accurately known by the authors.

6) A more striking fact will be seen in Fig. 4. The capacitance shown in Fig. 4 is a *model* of the junction capacitance of the tunnel diode. If the capacitance determined by the lumped parameter concept is C_0, then the admittance of the capacitance in the lumped parameter circuit concept is

$$Y_0 = j\omega C_0 \tag{1}$$

or

$$C_0 = \frac{Y_0}{j\omega} \cdot \tag{2}$$

If a short-circuited distributed parameter transmission line is attached to C_0 as shown in Fig. 4 and forms a short-circuited *uniform* loss-less transmission line of length L, then, the input admittance of the transmission line is[10]

[4] F. Sterzer and D. E. Nelson, "Tunnel-diode microwave oscillators," Proc. IRE, vol. 49, pp. 744–753; April, 1961.

[5] Received May 27, 1963.

[6] K. Ishii and C. C. Hoffins, "Extending tunnel diode operating frequency," *Electronics*, vol. 35, pp. 43–45; June, 1962.

[7] K. Ishii and C. C. Hoffins, "*X*-Band operation of *S*-band Esaki diodes," Proc. IRE (*Correspondence*), vol. 50, pp. 1698–1699; July, 1962.

[8] C. C. Hoffins and K. Ishii, "Microwave tunnel diode operation beyond cutoff frequency," Proc. IEEE (*Correspondence*), vol. 51, pp. 370–371; February, 1963.

[9] K. Ishii and C. C. Hoffins, "Oscillation frequency of microwave tunnel diodes," Proc. IEEE (*Correspondence*), vol. 51, pp. 485–486; March, 1963.

[10] S. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y.; 1943.

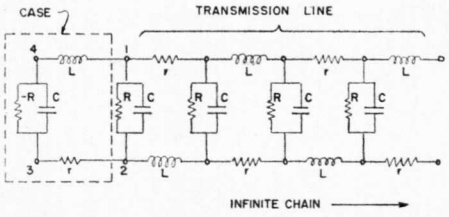


Fig. 3—An equivalent circuit of idealized theoretical transmission line attached to a microwave tunnel diode.

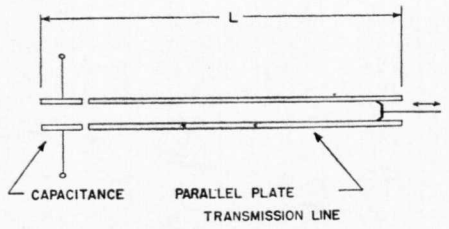


Fig. 4—A parallel plate transmission line attached to the junction capacitance of a microwave tunnel diode.

$$Y_0 = -jM \cot \beta L \tag{3}$$

where M is the characteristic admittance of the transmission line and β is the phase constant of the transmission line. In this configuration, if the equivalent capacitance across the tunnel diode junction in the distributed parameter circuit concept is C_d, then,

$$Y_0 = j\omega C_d \tag{4}$$

or

$$C_d = \frac{Y_0}{j\omega}. \tag{5}$$

Substituting (3) into (5)

$$C_d = -\frac{M}{\omega} \cot \beta L. \tag{6}$$

If the length and/or phase constant of the transmission line is adjusted so that

$$\beta L = (2n + 1)\frac{\pi}{2}, \quad n = 0, 1, 2, 3, \cdots, \tag{7}$$

then,

$$C_d = -\frac{M}{\omega} \cot (2n + 1)\frac{\pi}{2} = 0$$

or

$$C_d = 0. \tag{8}$$

Thus, the "equivalent" junction capacitance vanishes.

7) The above analysis is true only "in theory." Many practical design problems remain unsolved to realize the "theory." Finding a suitable design for a waveguide impedance transformer which will realize the above stated theory would be a key to solve these problems.

8) The authors appreciate McPhun's interesting remarks.

CORDELL C. HOFFINS
KORYU ISHII
Dept. of Elec. Engrg.
Marquette University
Milwaukee, Wis.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 51, NO. 9, SEPTEMBER, 1963

PRINTED IN THE U.S.A.

APPENDIX 3.

Proposals for a 2-way tunnel diode repeater using a 4-port circulator.

C.E.R.L. Laboratory Note No. RD/L/N 17/64, April 1964.

C.E.G.B. INTERNAL USE ONLY:

CENTRAL ELECTRICITY
RESEARCH LABORATORIES

LABORATORY NOTE No.RD/L/N 17/64.

(Job No. 41149)

PROPOSALS FOR A 2-WAY TUNNEL DIODE REPEATER USING A 4-PORT CIRCULATOR.

- by -

M.K. McPhun.

Approved by: D. F. Lucas.

Head of Division

Date: 14 April 1964.

SUMMARY

Work is in progress to develop non-demodulating repeaters, using tunnel diodes, for the 470 Mc/s frequency band. Detailed proposals are presented whereby such a repeater may be made; it would provide two-way amplification using a different centre frequency for each of the two directions. Common transmit and receive aerials, and a single 4-port circulator would be used.

Special narrow-band directional filters are required, and a method of making these filters is presented.

An experiment to verify some of the proposals is described. A crude directional filter was improvised and used, in conjunction with a 3-port circulator and a tunnel diode reflection amplifier, to make a stable one-way transmission amplifier having a gain of 29 dB at 530 Mc/s.

1. INTRODUCTION

Work on the tunnel diode radio repeater for the 470 Mc/s frequency band is directed towards making a non-demodulating repeater. Such a repeater consists essentially of an r.f. amplifier placed between two aerials; one aerial receives signals, the signal is passed through the amplifier, and is retransmitted by the other aerial.

A repeater must be able to amplify in both directions at once. The usual system uses a different frequency for each direction of transmission. The most elementary way of providing two-way amplification at a repeater station is to use two completely separate repeaters, one for each direction of transmission, each with its own pair of aerials. Although this is a practical solution it is evident that both capital and installation costs would be reduced if such items as the aerials could be common to both directions of transmission. This is the system which is often used with conventional repeaters.

When using tunnel diodes, we must consider carefully the stability of the circuit (McPhun 1963a), for the use of common transmit/receive aerials tends to aggravate the stability problem. The object of this note is to show that there are means whereby components common to both directions of transmission can be used, whilst maintaining stability of the whole system. The results of some confirmatory experimental work are given.

2. A STABLE ONE-WAY REPEATER

We have described in a previous report (McPhun 1963b) how we may make a transmission amplifier by combining a tunnel diode reflection amplifier with a 3-port circulator. To form a basic repeater (Fig. 1) from such an amplifier we simply substitute aerials for both the generator and the load (though parabolic aerials are shown in the diagram any kind of aerials may be used). Repeaters use high gain aerials and at U.H.F. such aerials usually have a narrow bandwidth. This fact makes the repeater of Fig. 1 very difficult to stabilize, and oscillations are liable to occur at some frequency remote from the resonant frequency of the aerials. We will examine how to make this into a practical one-way repeater.

2.1 Practical One-way Repeater

In a previous report (McPhun 963a) we discussed the factors affecting the stability of the simple repeater of Fig. 1, and derived five stability criteria which had to apply over a very wide bandwidth. The characteristics of narrow-band U.H.F. aerials at frequencies removed from resonance make these stability criteria extremely difficult to satisfy. Because an elegant solution to this stability problem requires a separate study which

we have not been able to undertake so far, we propose to use the following "belt and braces" solution; the circuit is designed so that the reflection amplifier sees the characteristic impedance Zo of the circuit at all frequencies, whilst on the other hand the circulator can see the tunnel diode reflection amplifier only over a very narrow band of frequencies for which the aerials present a good match to the circulator.

A practical way of providing the "belt and braces" solution is to interpose a directional filter (Wanselow, Tuttle 1959) between the circulator and the tunnel diode reflection amplifier.

2.1.1 The directional filter

The directional filter is essentially a transmission line component having three ports as shown diagramatically in Fig. 2a. The filter has a band-pass characteristic between ports A and B and a band-stop characteristic between ports A and C, as shown in Fig. 2b. However, it also has the following very important property when port C is terminated in a matched load (value Zo at all frequencies): we need only provide a matched termination to port B over the pass-band, to ensure that the impedance seen looking into port A is equal to Zo at all frequencies. For our purposes we are interested only in the band-pass characteristic so for the remainder of the report we shall assume port C to be terminated in a matched load, and not show it on subsequent diagrams. We shall now see how to use such a band-pass directional filter in the one-way repeater.

2.1.2 "Belt-and-braces" stability using a directional filter

Fig. 3 shows the repeater with a directional filter inserted between the reflection amplifier and the circulator. All the interconnecting lines are transmission lines with a characteristic impedance Zo.

Fig. 4 shows the pass-band of the filter with respect to the voltage standing wave ratio (v.s.w.r.) of the aerials. The filter pass-band is made narrower than that of the aerials, (the circulator pass-band would normally be wider than that of the aerials anyway). Now, providing that outside its pass-band the square of the filter attenuation exceeds the gain of the reflection amplifier, the repeater must be stable at all frequencies outside the pass-band of the directional filter.

We must now consider the risk of instability at the resonant frequency f_o, as instability at this frequency may occur if the aerials present a poor match. Here we must take into account not only the quality of the aerials used, but also factors outside our control such as the formation of ice, and birds alighting on the aerials. The next section will show how we may eliminate this cause of instability by using a perfect 4-port circulator.

2.2 The 4-Port Circulator used to Ensure Stability

The 4-port circulator should not be confused with the hybrid much used in line telecommunications. The circulator is shown in Fig. 5 where we see that it has four pairs of terminals or ports, each of these consisting of a transmission line entry. Power entering port A leaves at port B suffering no attenuation, and no power comes out of C or D. Similarly power entering at B leaves at C and none leaves at D or A and the same applies to ports D and A in cyclic order.

The basic method of making a tunnel diode amplifier using a 4-port circulator (Sie 1961) is shown in Fig. 6 where the circulator and all interconnecting transmission lines have the characteristic impedance Zo. Port D is terminated in a matched termination Zo so that any power that is reflected from the amplifier load back into port C will be directed out of port D and absorbed in this matched termination. Waves from the generator entering at port A are directed to the reflection amplifier on port B. They are amplified, re-enter at B and exit at C, and therefore pass to the load. Thus if the reflection amplifier is itself stable when connected to a characteristic impedance Zo, a perfect circulator gives us an absolutely stable amplifier. The reflection amplifier can see only the impedance Zo at all frequencies and any reflection from a mis-matched load is directed into the matched termination placed on port D of the circulator.

In practice the circulator is not perfect, that is, some small fraction of the power entering at C will appear at B and similarly some power entering at B will appear at A. Also there will be mis-matches at the ports. These two forms of imperfection will be worse outside the finite ranges of frequencies called the bandwidth of the circulator. Hence there is still a need for other stabilising techniques such as that already described in Section 2.1.

If we combine the stabilizing technique of Section 2.1 with a 4-port circulator, and if we use aerials in place of the generator and the load we have the circuit of the one-way repeater shown in Fig. 7. The directional filter has the characteristics of Fig. 4 already described, and thus the reflection amplifier sees the characteristic impedance Zo at all frequencies. The input and output aerials are in effect connected to the reflection amplifier only over a narrow bandwidth about f_o defined by the pass-band of the directional filter. Further, any reflections occurring at the output aerial at frequency f_o will be absorbed in the matched load Zo of Fig. 7. An additional advantage of this circuit is that any spurious signals entering the input aerial at frequencies outside the pass-band of the directional filter, will be

absorbed before reaching the reflection amplifier, and will not reach the output aerial.

The practical one-way repeater described above will be absolutely stable providing the circulator is perfect over a narrow band of frequencies about f . Thus it remains only to consider instabilities which may arise due to imperfections in the circulator at the resonant frequency. Worse case stability criteria for ensuring stability under these conditions have already been derived (McPhun 1963a) and will not be considered further in this note.

We shall now see how this one-way repeater may be extended into a two-way repeater.

3. THE PROPOSED TWO-WAY REPEATER

As already discussed, the most elementary method of two-way amplification at a repeater station is to use two separate one-way repeaters, each tuned to the appropriate resonant frequency, and each having its own aerials. As a first step of simplification we could use common aerials for the two repeaters providing the aerials are sufficiently wide-band to accommodate both resonant frequencies. We propose to go one step further and use a common 4-port circulator also, the arrangement being as shown in Fig. 8. In Fig. 8 the matched termination Zo of Fig. 7 is replaced by another directional filter with a different centre frequency, terminated in another reflection amplifier. Signals at frequency f_1 are amplified from left to right whilst signals at frequency f_2 are amplified from right to left, thus each aerial performs the dual functions of transmitting and receiving. Filter 1 is tuned to f_1 and presents a match at f_2; conversely filter 2 is tuned to f_2 and presents a match at f_1.

Consider the operation of this circuit whilst amplifying from left to right at frequency f_1. At this frequency we could replace filter 2 and its reflection amplifier by a matched termination; the circuit becomes as Fig. 7 but tuned to frequency f_1, and thus amplification takes place from left to right only. Any reflections occurring from the right-hand aerial will be absorbed in the matched termination and the amplifier will be stable. An exactly similar argument for f_2 shows that at this frequency the repeater will amplify from right to left and will again be stable.

It behoves us to see whether the components required for this two-way repeater are realisable; this we will do in the next section.

3.1 Practical Realisation of the Repeater

Our investigations show that the components required to realise this repeater are completely practicable. It is not necessary to present a detailed justification of this statement here but the following are some brief indications of the present state of the art.

At 470 Mc/s, 4-port circulators are available having a bandwidth of 5%; this is adequate for our purposes*. These circulators at present consist of two cascaded 3-port circulators built as a single unit. For aerials at 470 Mc/s we propose to use stacked and bayed Yagi arrays. These have 20 dB gain and just enough bandwidth to accommodate the two frequencies f_1 and f_2.

We have made tunnel diode reflection amplifiers with a gain-bandwidth product of 750 Mc/s, more than adequate for use in the present circuit. These will be described in a subsequent report.

It is worth while noting here that at present all the above components with the exception of the tunnel diodes themselves, can be made more compact, cheaper, and with a better performance, at frequencies higher than 470 Mc/s.

It remains to examine the requirements for the special directional filters.

3.1.1 Proposed directional filters

Directional filters can conventionally be made as a single component in strip-line (Wanselow, Tuttle 1959). However, for the narrow pass-band required this type of filter has too much insertion loss in the pass-band at a centre frequency of 470 Mc/s. At higher frequencies a waveguide construction will give a satisfactory component with low insertion loss in the pass-band, but at 470 Mc/s such a filter measures 2 ft. 6 in. diameter by 3 ft. 4 in. high and this is a little ludicrous when we consider that the tunnel diode is smaller than a match head.

We propose to make directional filters by the method shown in Fig. 9. This filter uses a 3-dB hybrid and two conventional band-stop filters, each terminated in matched loads. Referring to Fig. 9 the 3-dB hybrid has the following properties. Power entering at port A splits equally and leaves by ports B and C there being a 90^0 phase shift between the signals leaving at ports B and C; no power appears at D. Similarly, power entering at B leaves by ports A and D and no power leaves at C, and the same applies in cyclic order to power entering at ports C and D.

The band-stop filters are conventional in that they present a complete rejection of the input signal in the stop band. They are terminated in matched loads so that when looking into the filters we see a reflection coefficient Γ_{in} as shown in Fig. 10. Thus the input reflection coefficient is zero at all frequencies except within the stop-band

---

* Since writing this, 4-port circulators with a bandwidth of 20% have become available at 470 Mc/s.

of the filter where it is unity, i.e. all power is reflected. Also shown in Fig. 10 is the characteristic of the complete directional filter of Fig. 9, which we will consider separately for the pass-band and the stop-band.

Firstly, consider the action of the directional filter within its pass-band. In Fig. 9, power P entering at A splits equally with a 90^{o} phase shift and exits at B and C. All this power is reflected by the band-stop filters and re-enters the hybrid at B and C. This reflected power combines in-phase at D and anti-phase at A, so that all the power P originally entering at A leaves port D suffering no attenuation. As no fraction of power P is reflected at A, then port A must represent a matched input.

Secondly, consider the operation of the directional filter within its stop-band. Again power P entering at A splits equally and exits at B and C. But here the power passes through the band-stop filters to be absorbed in the two matched loads Zo, so that port A again represents a matched input and furthermore no power leaves port D.

The circuit of Fig. 9 thus acts as a true directional filter and whilst possessing an overall band-pass characteristic it also presents a match at all frequencies. The hybrid is a standard component readily available, so that the realisability rests with the required performance of the band-stop filters. We have insufficient effort available at present to undertake the design and construction of the filters at C.E.R.L., but inquiries from two manufacturers have shown that our specification for these filters is not unreasonable.

The hybrid, filters and a reflection amplifier, could be made as a single compact integrated unit quite cheaply if the number of repeaters required justified it.

In order to prove some of the above proposals we have performed the experiments described in the following section.

4. AN EXPERIMENTAL TUNNEL DIODE CIRCULATOR AMPLIFIER

The experiments described here are not meant to be elegant; they were performed as quickly as possible using standard components. The object was two-fold: firstly to improvise a crude directional filter using two very simple band-stop filters made from standard co-axial components, and secondly to use this improvised directional filter to make a one-way circulator amplifier using a 3-port circulator and a reflection amplifier already in our possession.

4.1 Experimental Directional Filter

This directional filter has a schematic form already shown in Fig. 9. The hybrid used is a Narda type 3031 intended to cover a frequency band from 460 Mc/s to 950 Mc/s.

The schematic diagram of the band-pass filters

is shown in Fig. 11a. At the resonant frequency the transmission line is a quarter wave-length long and thus the 50-ohm resistor appears to be short-circuited. The input impedance to the filter approaches 50 ohms at D.C. and also when the transmission line is half-wave length long. (spurious responses will be obtained whenever the transmission line is an odd number of quarter wave-lengths long). Fig. 11b shows the somewhat primitive construction of these filters. They were made from a standard coaxial T-piece, one arm being the input, the other two arms being terminated, respectively, in a matched 50-ohm coaxial load and a short length of coaxial cable.

The complete directional filter gave the results shown in Fig. 12. It is seen that the bandwidth is very wide, an attenuation of 3 dB being obtained at frequencies of 660 Mc/s and 370 Mc/s. However, the filter has zero attenuation at the resonant frequency of 515 Mc/s, and also has the required match at all frequencies; the v.s.w.r. was less than 1.2 over the whole frequency range from 40 Mc/s to 800 Mc/s.

4.2 The Reflection Amplifier

We have been experimenting with reflection amplifiers for some time, and details of these will be published in a subsquent report. However, we are interested in the gain/frequency characteristic of the amplifier used, and this is shown in Fig. 13. We should note that this gain/frequency characteristic will only be obtained when the amplifier "sees" a 50-ohm load at all frequencies. If the amplifier sees loads of other impedances or loads which vary with frequency, we may obtain other gains or the amplifier may oscillate (McPhun 1963a). When connected to a 50-ohm load the reflection gain at the resonant frequency of 500 Mc/s is 16.3 dB, and the 3-dB bandwidth is 34 Mc/s.

4.3 The Complete Amplifier - Results

The 3-port circulator used for this experiment has a very complex impedance vs. frequency characteristic. When terminated in matched loads on two ports the third port presents a match for a small frequency band about 500 Mc/s, and a large mis-match at many other frequencies. Hence the reflection amplifier oscillated when connected directly to the circulator.

The schematic diagram of the amplifier Fig. 14 is similar to that previously shown in Fig. 3, but here we use a matched resistive generator and load. The actual arrangement of the amplifier is shown in Fig. 15.

The completed amplifier was stable, stability being checked with a tunnel diode stability tester by examining the current versus voltage characteristic of the tunnel diode (McPhun 1962).

The amplifier was also stable with either the input or the output short-circuited, and with a variety of mis-matches on the input and output terminals.

The gain vs. frequency characteristic of the amplifier is shown in Fig. 16. This has two peaks, at 490 and 530 Mc/s respectively, these being due to the shortcomings of the filters. The circulator and the tunnel diode reflection amplifier were each tuned to a 500 Mc/s centre frequency, and the directional filter to 515 Mc/s, and no attempt was made to tune the complete amplifier after assembling the individual parts. Also, neither the circulator nor the directional filter presented a very good match at the resonant frequency. We have been able to explain the peaks quantitatively, but it is not necessary to reproduce this work here. We achieved the main objects of this experiment which were to make a primitive directional filter by the method proposed in Section 3.1.1 and to use it to make a stable circulator amplifier.

5. CONCLUSIONS

We have proposed a method for making a stable 2-way tunnel diode repeater for amplifying a different frequency in each direction of transmission. This repeater would use common transmit and receive aerials for each direction of transmission and a common 4-port circulator.

Directional filters, which are required anyway at present to ensure stability, are used to separate the two frequencies.

Investigations show that all the components required are realisable and available in a reasonable time. However, we have found that all the parts, i.e. the circulator, filters, and the aerials would be more compact, cheaper, and have better performance, if higher frequencies than 470 Mc/s were used. As there is an upper limit to the frequency at which we can use tunnel diodes, further work is needed to establish an optimum frequency for the use of tunnel diode repeaters.

For the immediate requirements at 470 Mc/s, conventional directional filters are not satisfactory. We have described a method for making directional filters at this frequency and have verified experimentally that this can be done. We have made a tunnel diode amplifier using such an experimental directional filter in conjunction with a 3-port circulator. This amplifier was stable and gave 29 dB gain at 530 Mc/s.

These results are very encouraging and give us every reason to believe that a 2-way repeater made to these proposals will perform to our predictions.

The proposed method of constructing a repeater should be adopted for the job in hand at 470 Mc/s.

6. REFERENCES.

McPhun, M.K. 1962, C.E.R.L. Note No. RD/L/N 92/62.

McPhun, M.K. 1963a, C.E.R.L. Note No. RD/L/N 105/63.

McPhun, M.K. 1963b, C.E.R.L, Memorandum No. RD/L/M 43.

Sie, J.J. 1961, "Application of Tunnel Diode Amplifiers"; 3, (Nerem Record).

Wanselow, R.D. and Tuttle, L.P. 1959, Practical Design of Strip-transmission Line Half-wavelength Resonator Directional Filter, Trans.I.R.E., M.T.T., p.168.

Drgs. No. RL.2.1.801 to 808 attached.

Copies to:

Member for Research
Chief Scientist
Chief Research Officer
Directors (R&D)
Technical Planning Officer (R&D)
Mr J.M. Ward
Transmission Plant Design Engineer
Mr S.A. Clarke
Mr L. Csuros
Mr A.U. Dowell
Mr C.W. Mott
Mr J.H. Naylor
Regional R & D Officers
Chief Engineers, Area Boards
Telecommunication Engineers
Mr A.J. Davidson (N.S.H.E.B.)
Mr S.C. Geerke (S.S.E.B.)
Tech. Sub-Committee N.F.P.I.(6)
Information Section
Librarian (BNL)
Librarian (CEGB)
Librarian (MEL)
Registry.

DAR.

Text cut off in original

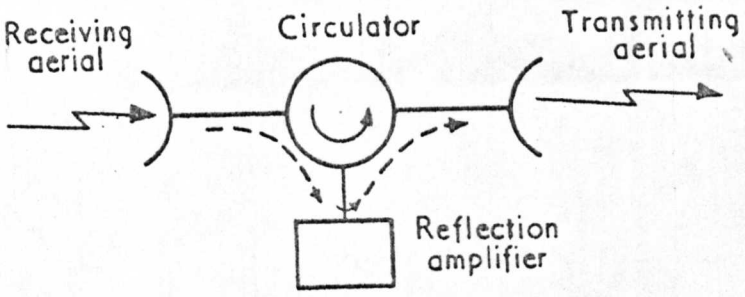


FIG.1 BASIC REPEATER

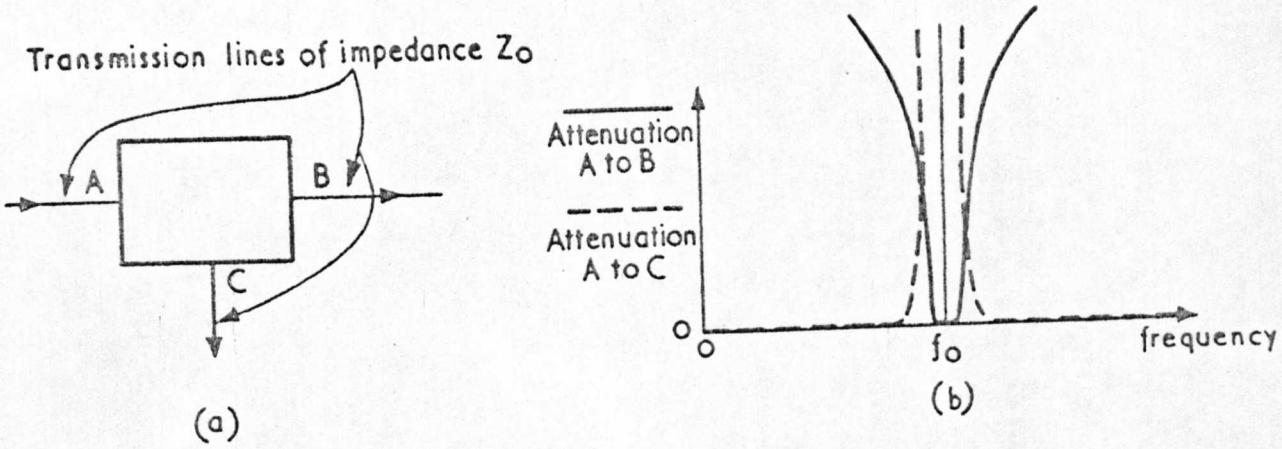


FIG. 2 SCHEMATIC (a) AND PERFORMANCE (b) OF DIRECTIONAL FILTER

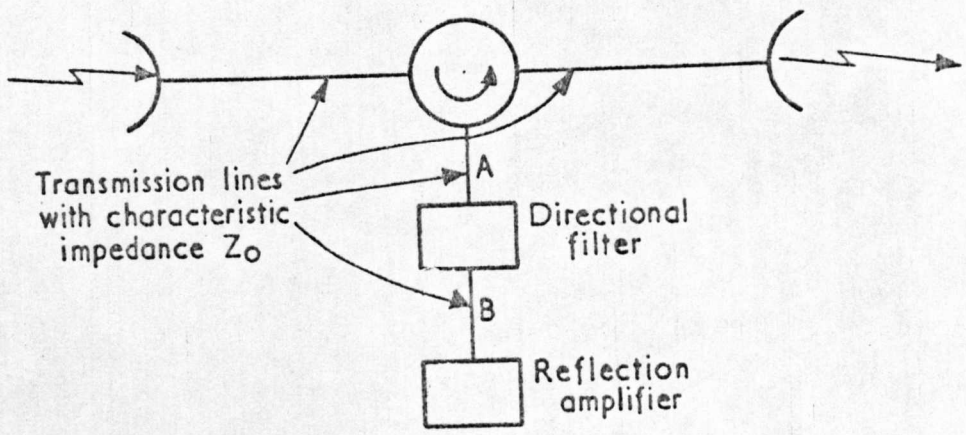


FIG. 3 DIRECTIONAL FILTER USED TO ATTAIN STABILITY

DRAWN BY: M.K.M. / J.B.C. DATE: 10·2·64 DRG No. RL 3·1·801

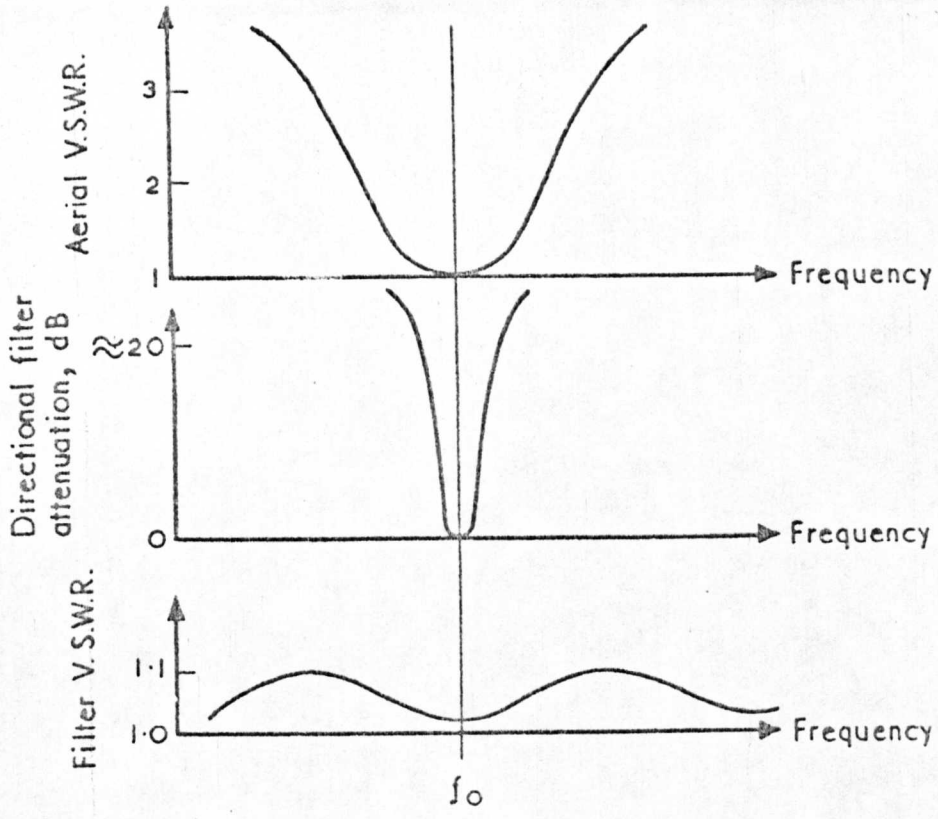


FIG.4 RELATIVE RESPONSES OF AERIALS AND DIRECTIONAL FILTER

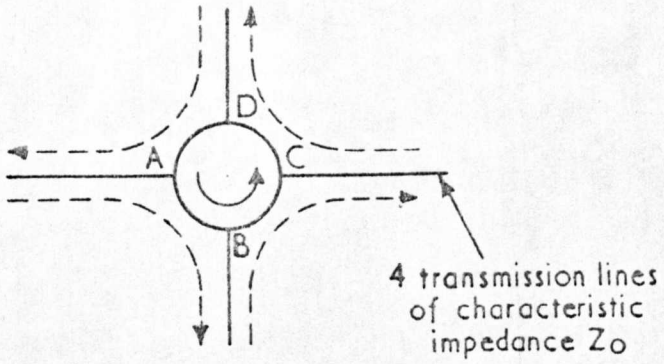


FIG.5 THE 4-PORT CIRCULAR

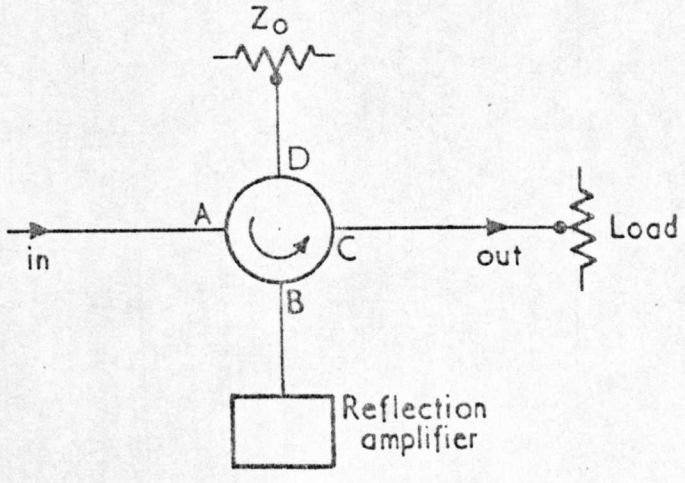


FIG.6 BASIC AMPLIFIER USING A 4-PORT CIRCULATOR

DRAWN BY: M.K.M./J.B.C. DATE: 11·2·64 DRG. No. RL.2·1·802

LR 1839

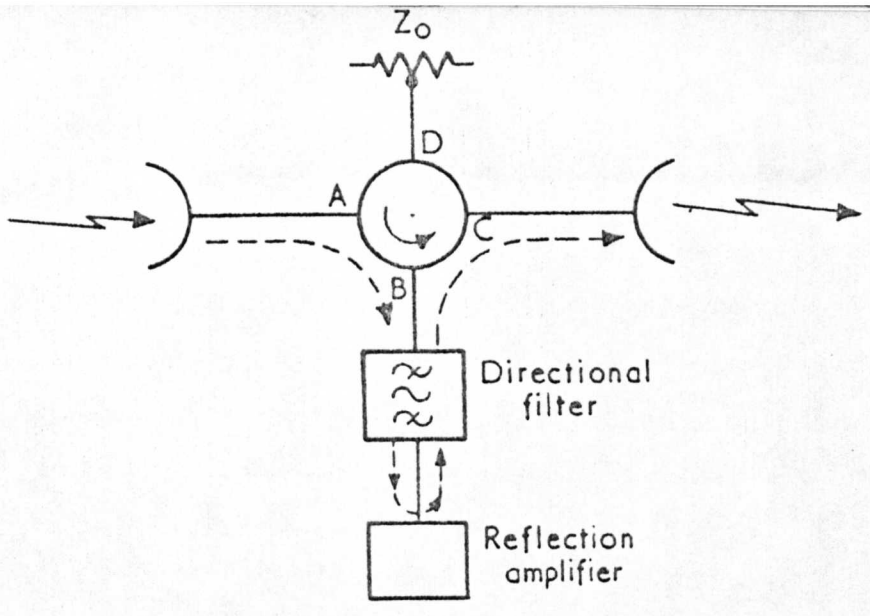


FIG.7 PRACTICAL ONE-WAY REPEATER

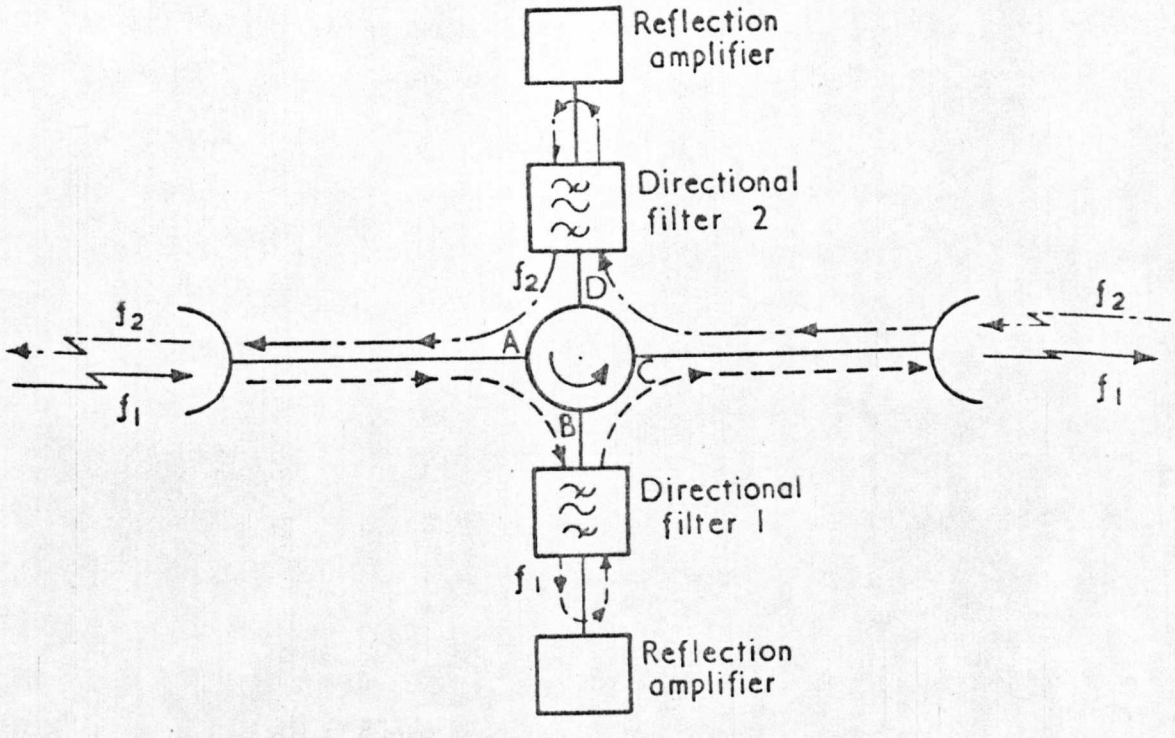


FIG.8 PROPOSED TWO-WAY REPEATER

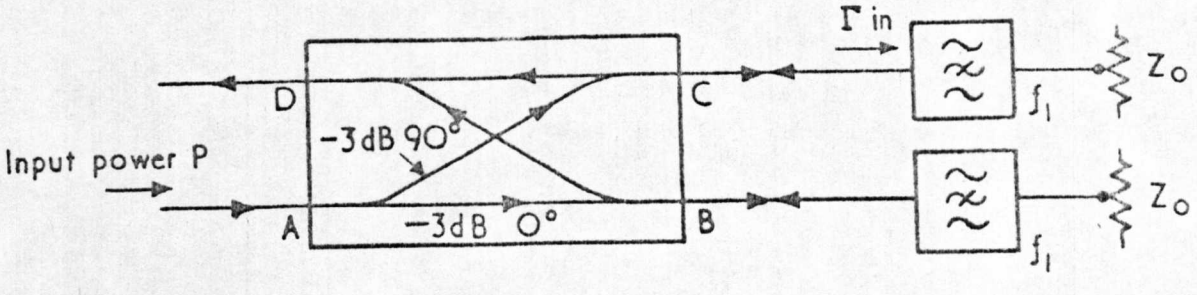


FIG.9 PROPOSED DIRECTIONAL FILTER

DRAWN BY: M.K.M./J.B.C. DATE: 13.2.64 DRG. No. RL 2-1-803

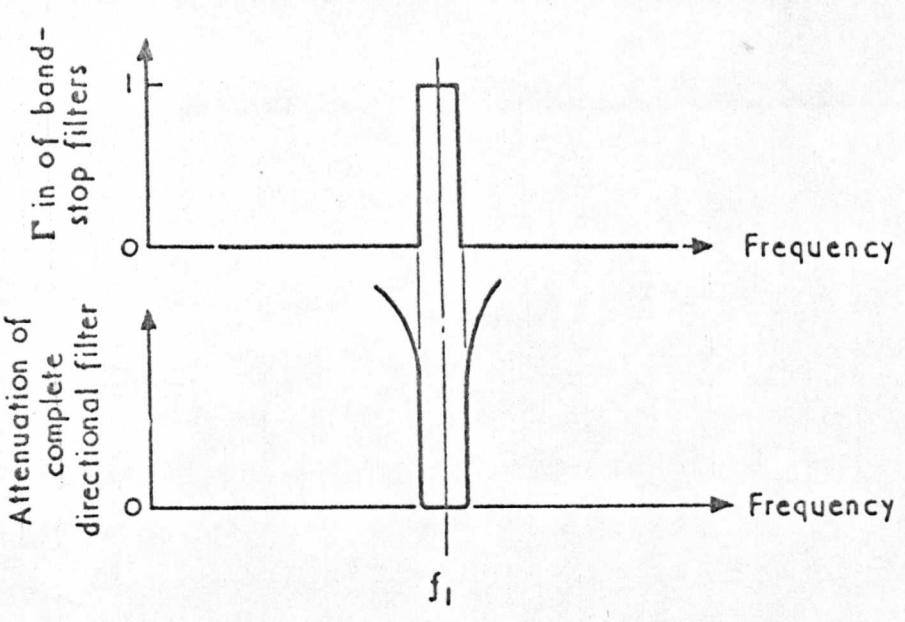
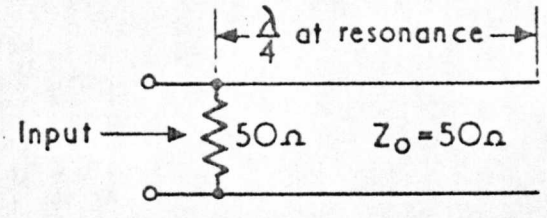
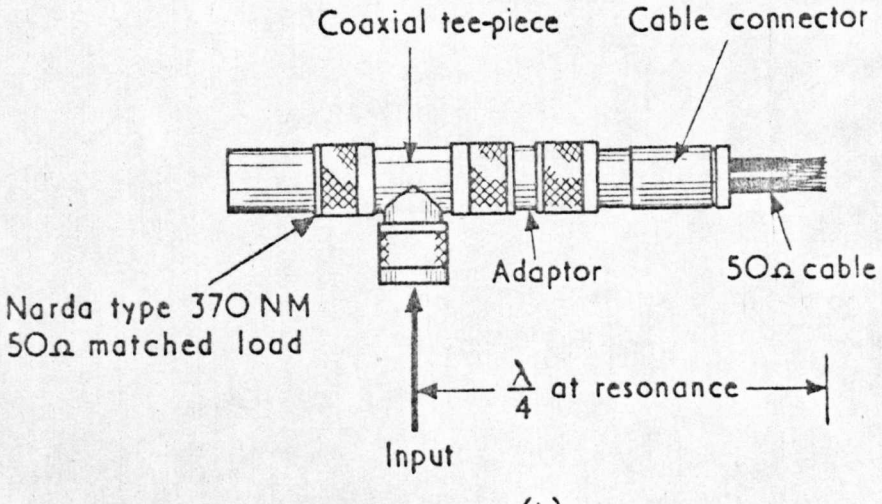


FIG.10 CHARACTERISTICS OF PROPOSED DIRECTIONAL FILTER



(a)



(b)

FIG.11 SCHEMATIC (a) AND ACTUAL ARRANGEMENT (b) OF PRIMITIVE BAND-STOP FILTERS

DRAWN BY: M.K.M. / J.B.C. DATE: 14·2·64 DRG No. RL 2-1-804

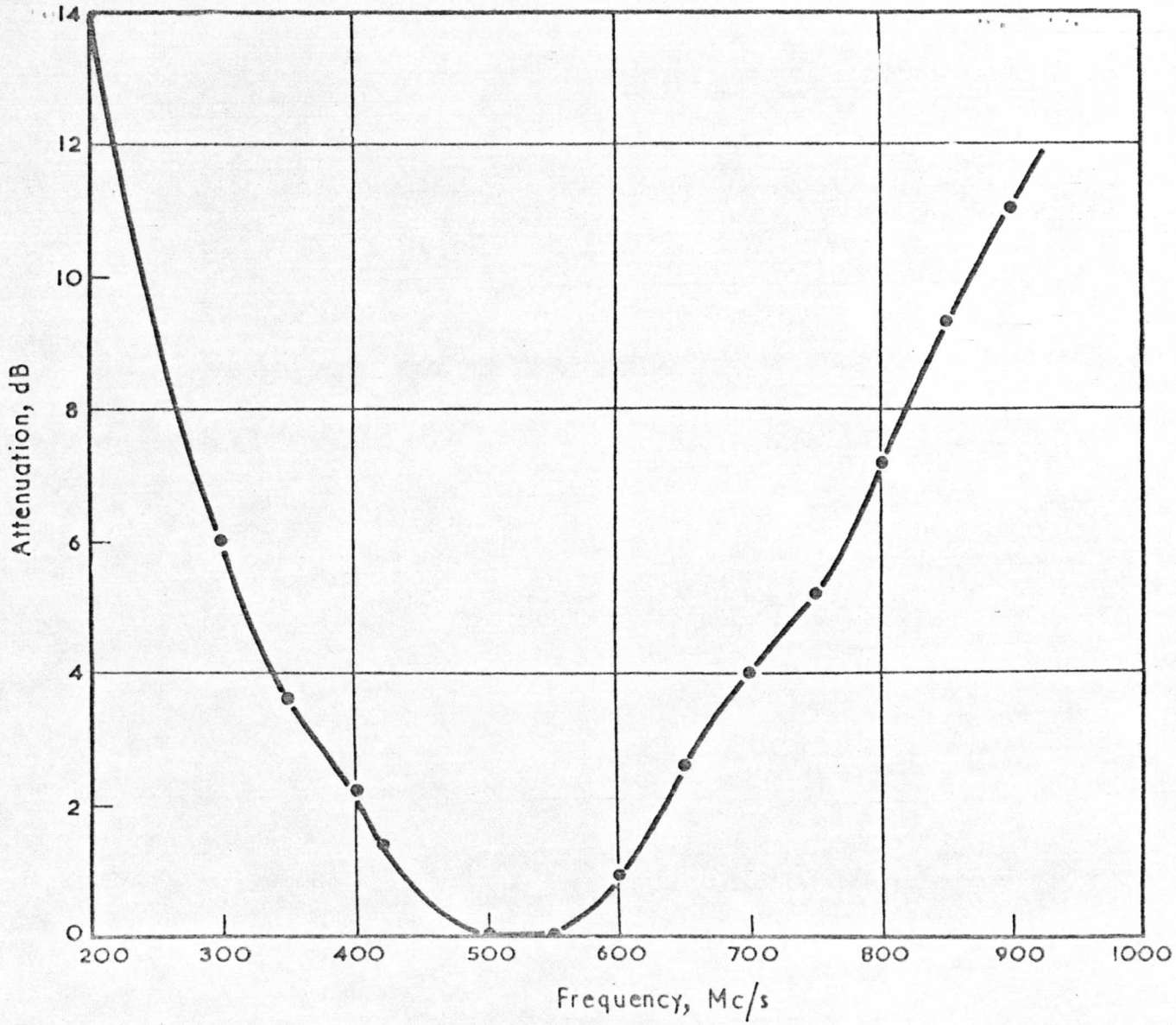


FIG.12 ATTENUATION VERSUS FREQUENCY FOR EXPERIMENTAL DIRECTIONAL FILTER

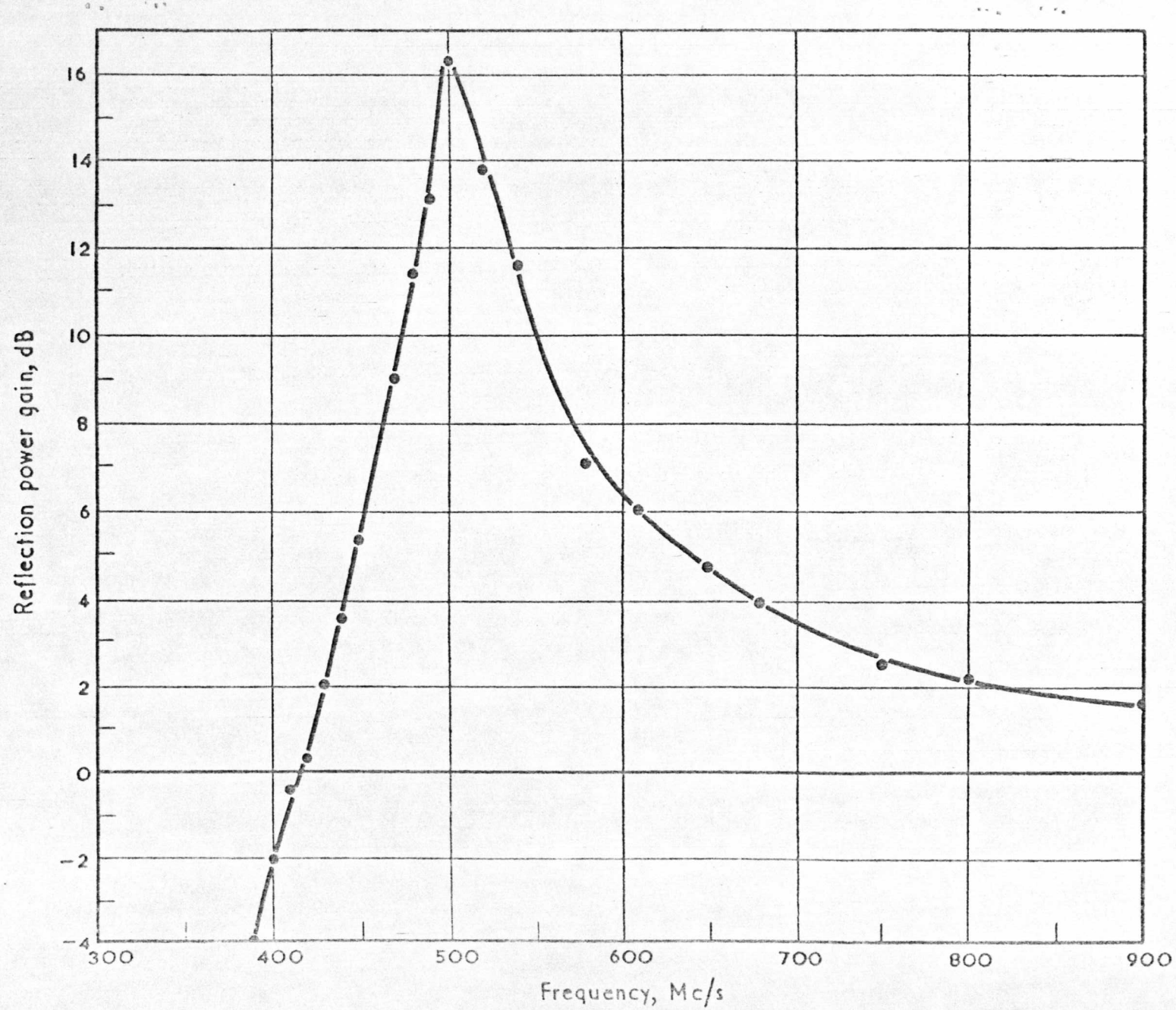


FIG.13 REFLECTION POWER GAIN VERSUS FREQUENCY FOR THE REFLECTION AMPLIFIER

DRAWN BY: M.K.M./J.B.C DATE: 13·2·64 DRG. No. RL. 2·1·806

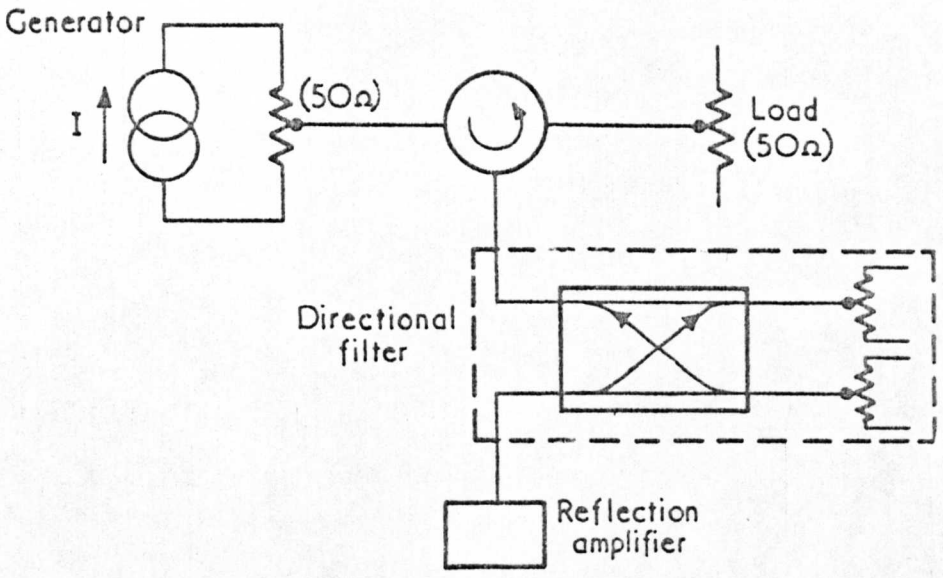


FIG.14 SCHEMATIC OF EXPERIMENTAL TUNNEL DIODE CIRCULATOR AMPLIFIER

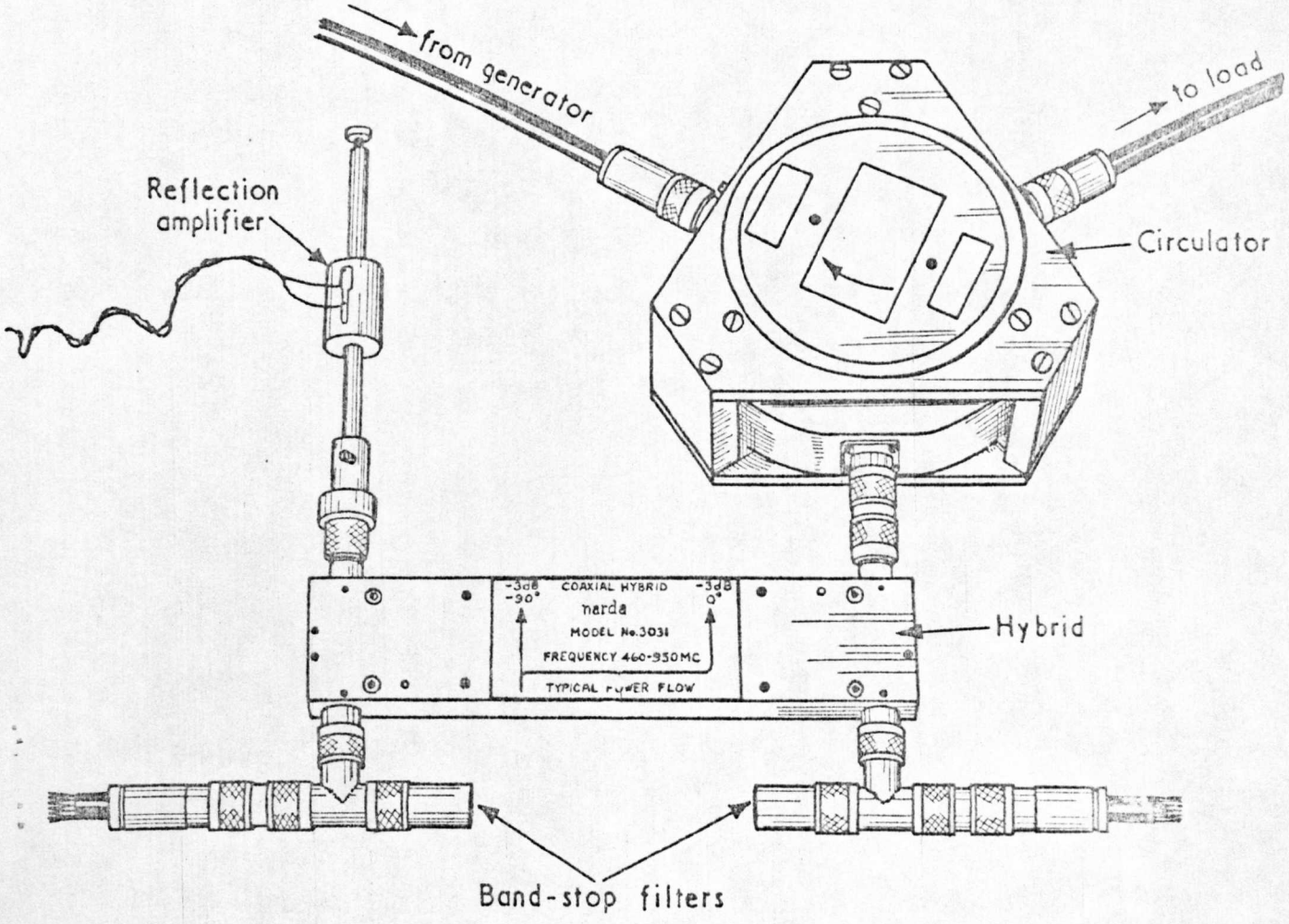


FIG.15 ACTUAL ARRANGEMENT OF EXPERIMENTAL TUNNEL DIODE CIRCULATOR AMPLIFIER

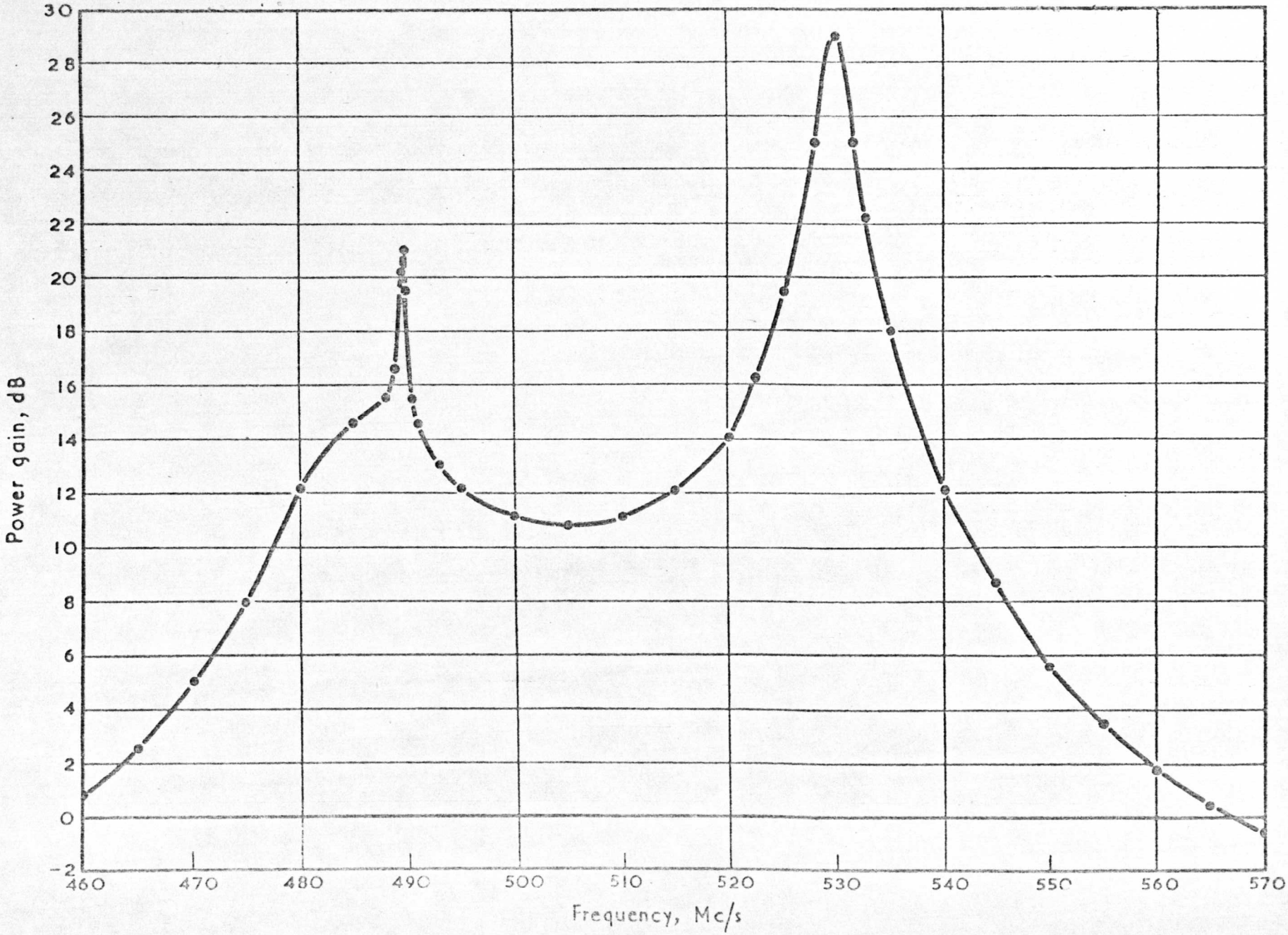


FIG.16 POWER GAIN VERSUS FREQUENCY FOR THE CIRCULATOR AMPLIFIER SHOWN IN FIG. 15

DRAWN BY: M.K.M./J.B.C. DATE: 13·2·64 DRG. No. RL. 2·1·803

LK 1339

APPENDIX 4.

Practical stability criterion for tunnel-diode circuits.

Electronics Letters 1, 9, pp 167-168, (Aug. 1965).

Reprinted from 'Electronics Letters', August, 1965

PRACTICAL STABILITY CRITERION FOR TUNNEL-DIODE CIRCUITS

A stability criterion for tunnel-diode circuits is described, which requires only plots of modulus of reflection coefficient versus frequency. The criterion is derived intuitively.

The designer of tunnel-diode amplifiers needs a method of predicting whether his circuit will oscillate or not. So far, the only stability criterion suitable for practical application to tunnel-diode amplifiers in general is that published by Meinke.[1] To use Meinke's criterion, one plots on a Smith chart the circuit impedance 'seen' by the negative resistance $-R$ of the tunnel-diode junction; if, over the frequency range d.c. to diode cutoff, the impedance locus does not encircle the centre of the chart, the circuit is stable: otherwise it is unstable. This criterion has two disadvantages: first, the circuit behaviour must be examined at the terminals of the junction negative resistance, and any extension of the criterion to allow the use of other terminals makes its use very cumbersome; secondly, both amplitude and phase information must be plotted. The following criterion has neither disadvantage.

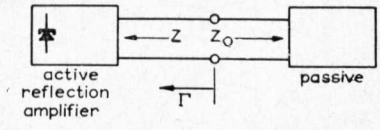


Fig. 1 *Circuit divided into active and passive parts*

Statement of criterion

The stability criterion may be stated as follows. Divide the circuit into any two parts, one part (the reflection amplifier) containing the junction negative resistance and the other part containing only passive circuit elements (Fig. 1). For $R = R_0$ (the nominal value of junction negative resistance), compute the modulus of reflection coefficient $|\Gamma|$ of the reflection amplifier as seen by the passive circuit, and plot this against frequency from d.c.

to diode cutoff. Increase the magnitude of R slightly and replot the curve. If the increase in negative resistance results in an increase of reflection coefficient, the amplifier is unstable: otherwise it is stable. Figs. 2*a* and *b* show simple examples of the curves obtained for unstable and stable amplifiers. We may derive this criterion intuitively as follows.

Intuitive derivation

Curves of magnitude and phase of reflection coefficient Γ for active circuits may be plotted as $\Gamma' = -1/\Gamma$ on the negative Smith chart.[2] This chart is used in Fig. 3; $|\Gamma|$ is unity at the rim and rises to infinity at the centre.

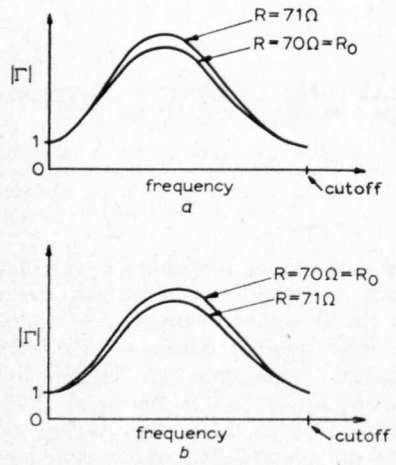


Fig. 2 *Modulus of reflection coefficient versus frequency*

a Unstable circuit
b Stable circuit

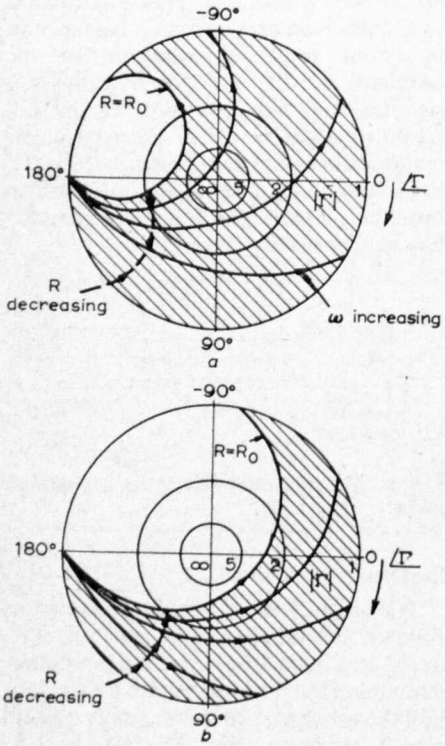


Fig. 3 *Negative Smith-chart plots of reflection coefficient*

As R is decreased from infinity to its final value R_0
a Unstable circuit
b Stable circuit

We may calculate a frequency-response curve using the small-signal equivalent circuit of the tunnel diode. This curve will give no information about stability, except for the one special unstable case where at some frequency $|\Gamma|$ equals infinity; i.e. the curve passes exactly through the centre of the chart. As the circuit stability depends on whether the active element is voltage- or current-controlled, it is necessary to take the nonlinearity of the device into account. The tunnel diode is voltage-controlled; hence, as the bias voltage is increased from zero to the operating point, the magnitude of its negative resistance $-R$ must decrease from infinity to its final value R_0.

In Fig. 3 families of reflection-coefficient curves are plotted for a simple case. The parameter is R (a positive number), which is decreased from infinity down to R_0, causing the curves to trace out the shaded areas shown.

The unstable amplifier is distinguished by the fact that, for some particular value of the negative resistance within that range, the curve passes through the point $|\Gamma| = \infty$. It is clear from the foregoing argument that Fig. 3*a* represents an unstable circuit and Fig. 3*b* represents a stable circuit.

If we plot the curves of $|\Gamma|$ versus frequency for $R = R_0$ as in Fig. 2, it is quite possible for both the stable and unstable cases to have the same value of peak gain. But inspection of Fig. 3 shows that any increase of R will result in an increase of $|\Gamma|$ for an unstable circuit (Fig. 2*a*), whereas for a stable circuit $|\Gamma|$ will decrease (Fig. 2*b*). Therefore the curves of Fig. 2 contain all the information required to indicate stability.

Discussion

For clarity we have shown a particularly simple example in Fig. 3, where the $R = R_0$ locus encircles the centre of the chart for the stable case (this should not be confused with Meinke's impedance-plot stability criterion). We are not sure at present whether this condition will always apply to more complicated amplifiers.

$|\Gamma|$ is given by

$$|\Gamma| = \left|\frac{Z - Z_0}{Z + Z_0}\right|$$

where both Z and Z_0 may be complex. However, the power-reflection coefficient[3] $|S|^2$ may equally well be used in place of $|\Gamma|$. Here

$$|S|^2 = \left|\frac{Z - Z_0^*}{Z + Z_0}\right|^2$$

and it can be seen that both $|\Gamma|$ and $|S|^2$ tend to infinity for the same condition $Z = -Z_0$. We have found it most convenient to plot reflection gain in decibels:

$$S^2\,\mathrm{dB} = 20 \log_{10} \left|\frac{Z - Z_0^*}{Z + Z_0}\right|$$

which also gives the transmission gain of a circulator-coupled amplifier.

It may be essential to include any nonreciprocal elements in the passive block of Fig. 1, but this has still to be investigated. However, for circulator-coupled amplifiers it is most convenient to take a reference plane at the junction of the reflection amplifier and the circulator.

We have tested the criterion by applying it to various circuit configurations and comparing the results with other stability criteria such as Meinke's. We tested freedom to select any reference plane by deliberately choosing most unlikely reference terminals. The criterion passed all these tests. So far we have not proved the criterion in a rigorous theoretical way, but we are now endeavouring to do this.

M. K. MCPHUN
Mullard Research Laboratories
Redhill, Surrey
England

28th July 1965

References

1 MEINKE, H. M.: 'Antenna with tunnel diode', ASTIA Report AD 404714
2 KYHL, R. L.: 'Plotting impedances with negative resistive components', *IRE Trans.*, 1960, **MTT-8**, p. 377
3 KUROKAWA, K.: 'Power waves and the scattering matrix', *IEEE Trans.*, 1965, **MTT-13**, p. 194

Adlard & Son, Ltd., Bartholomew Press, Dorking

APPENDIX 5.

Practical stability criterion for tunnel-diode circuits.

Electronics Letters, 1, 10, p (Dec. 1965).

Reprinted from *Electronics Letters*, Vol. 1, No. 10, December 1965

PRACTICAL STABILITY CRITERION FOR TUNNEL-DIODE CIRCUITS

In a reply to Bandler's criticisms of the stability criterion, it is confirmed that the criterion includes both bias and stabilising circuits. No exceptions to the criterion have been proved so far.

In a previous letter,[1] I described a stability criterion requiring only plots of modulus of reflection coefficient versus frequency. Bandler[2] has since suggested that the criterion may not be generally valid.

Although Bandler suggests that the criterion does not include bias and stabilising circuits, both have, in fact, always been included in my work with this criterion. They may be included in either Z or Z_0 of my Fig. 1. Perhaps this confusion results from Bandler's Fig. 1, where he assumes only a lossless network between the tunnel diode and resistive load; I have not made this assumption. Thus, in his paragraph 4, $|\Gamma(j\omega)|^2$ is only independent of the reference plane chosen for his assumption of lossless networks. One 'most unlikely' reference plane is a plane that is isolated from the rest of the circuit by a large amount of loss; the criterion has been found to work even under these conditions.

In his penultimate paragraph, Bandler states 'The d.c. gain in McPhun's example could not be unity when the bias circuit is accounted for; therefore . . .' Fig. 1 shows a circuit which gives gain curves similar to my earlier Fig. 2, with a d.c. gain of unity.[1]

We now consider Bandler's Fig. 2*c*, which showed an interesting impedance locus that would invalidate my criterion; the circuit and locus are reproduced in Fig. 2 for clarity.

For $B > G$, Bandler shows that, as R is increased, the locus in quadrant 2 of Fig. 2*b* must move towards the origin, increasing $|\Gamma|$; yet this locus represents a stable amplifier. This locus is closely analagous to Nyquist's classic conditionally stable feedback amplifier. Let us consider what happens to $|\Gamma|$ as the tunnel-diode bias is increased through the negative-resistance region, remembering that $|\Gamma|$ becomes infinite when

$$Z_L(j\omega) + Z_C(j\omega) = 0$$

As the bias voltage is increased (R reduced) to a value giving the locus of Fig. 2*b*, $|\Gamma|$ must pass through infinity twice, with an unstable region when the locus of Fig. 2*b* does not encircle the origin. Further increase of bias reduces R (approaching the valley region), and again the amplifier becomes unstable. Thus this stable condition could only be reached by first passing through an unstable condition, and in practice tunnel-diode amplifiers do not behave like this.

We are led, therefore, to seek a theoretical reason for excluding a locus of this shape. Such a reason might be provided by the constraints imposed upon Z_C. First, Z_C must be positive real. Secondly, Z_C must have one more left half-p plane zero than it has poles, as its first element must be a series inductance. It seems quite likely that the locus of Fig. 2*b* could not be obtained under these constraints.

To conclude, it appears that no invalidity of the criterion has been proved so far, and a theoretical derivation, although difficult, would be very worth while. I thank Mr. Bandler for his contribution.

M. K. McPHUN *29th November 1965*

Mullard Research Laboratories, Redhill, Surrey, England

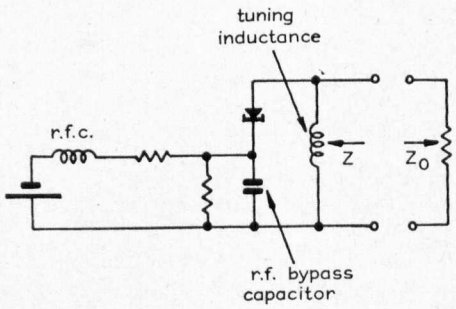
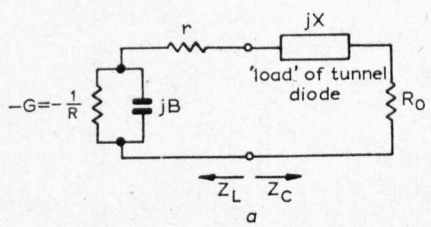


Fig. 1 *Tunnel-diode amplifier*



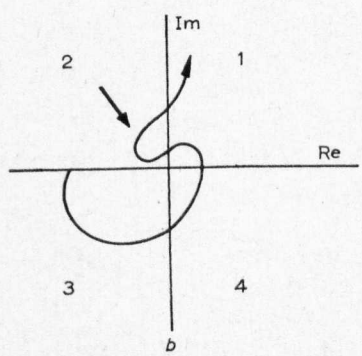


Fig. 2
a Amplifier representation
b Locus in $\{Z_L(j\omega) + Z_0(j\omega)\}$ plane

References

1 McPHUN, M. K.: 'Practical stability criterion for tunnel diode circuits', *Electronics Letters*, 1965, **1**, p. 167
2 BANDLER, J. W.: 'Practical stability criterion for tunnel diode circuits', *ibid.*, 1965, **1**, p. 262

APPENDIX 6.

Measurement of negative resistance using a Z-g Diagraph.

Proceedings I.E.E.E. 54, 6, pp 910-911, (June 1966).

Measurement of Negative Resistance Using a Z-g Diagraph

I. Introduction

The reflection gain of reflection amplifiers may conveniently be measured using a good directional coupler to separate the incident and reflected waves. However, when phase information is required, one must usually resort to laborious slotted line measurements.

Complex impedance and admittance (also two-port immittances) may be measured accurately and quickly using a Z-g Diagraph,[1] the results being obtained as plots on a Smith chart. We will see that only a simple modification of the measuring technique is required to measure active devices, negative resistance and reactance being obtained directly on the negative Smith chart.[2,3]

II. Standard Measurement Procedure

The Diagraph operates on the principle of comparing the magnitude and phase of the reflected voltage from the "unknown" with that from a standard (usually a short circuit), the incident voltage being the same in each case. As shown in Fig. 1, power incident from the signal generator at the input splits equally between the reference line and the measurement line, and is reflected, respectively, from the reference and the unknown terminations. The directional coupler in the measurement line gives a voltage proportional to the reflection coefficient Γ_x of the unknown, while that in the reference line gives a voltage proportional to the reflection coefficient unity. Thus $|\Gamma_x|$ can be obtained from the ratio of these two voltages, while the phase angle between them equals $\angle\Gamma_x$. A superheterodyne receiver is used for both signals, the local oscillator being common to both. $\angle\Gamma_x$ is measured at the intermediate frequency and indicated by a null on a meter M. The amplitude of the reference voltage is shown on a meter, while that reflected from the unknown deflects a light spot galvanometer. The instrument is constructed mechanically, such that the light spot indicates impedance directly on a Smith chart mounted on the front of the instrument.

The standard measurement procedure of interest here is as follows.

a) Standardize
   1) Place a short circuit on both reference and unknown ports
   2) Set signal generator to give light spot at rim of chart
   3) Set reference meter to read 100
   4) Set phase adjust to give a null on M when chart is set to zero phase angle (zero ohms for impedance, ∞ for admittance measurements).

b) Measure
   5) Connect unknown, to "unknown" port
   6) Obtain phase balance by rotating chart to again give a null on M
   7) Mark position of light spot on the Smith chart
   8) After changing frequency it is only necessary to adjust the signal generator output to again set the reference meter on 100, and repeat steps 6) and 7) without disturbing the connections to the unknown.

III. Measurement of Impedance and Admittance with a Negative Real Part

Negative impedances give reflection coefficients greater than unity. They may be plotted conveniently on a negative Smith chart by employing the transformation Γ chart $= 1/\Gamma_x$.[2,3,4]

The essence of the measuring technique used here is the interchange of the reference and unknown ports. The item to be measured (having negative resistance) is placed on the reference port, causing the reference meter to deflect off scale. However if the signal generator output is now reduced to bring the reference meter back to 100, the light spot deflection is reduced to a magnitude $1/|\Gamma_x|$. The phase angle given by the bridge at balance will now be the negative of that obtained for conventional connection, and the direction of rotation about the Smith chart will be reversed. This is exactly the condition required for plotting on the negative[2] or double[3] Smith chart.

The measurement procedure now is as follows.

a) Standardize
   1), 2), 3), 4) as described in Section II.

b) Measure
   5) Connect the unknown to the "reference" port
   6) Reduce signal generator output to again set the reference meter to read 100

Manuscript received February 23, 1966. The work reported herein was performed at the Central Electricity Research Laboratories, and is published by permission of the Central Electricity Generating Board.

[1] R. Eichacker, "The Z-g Diagraph," Rhode and Schwarz Mitteilungen, 1952, p. 78 (in German). See also instrument handbooks for types ZDU and ZDD Diagraphs.

[2] B. Rosin, "Transformation of impedances having a negative real part and the stability of negative resistance devices," *Proc. IRE (Correspondence)*, vol. 48, p. 1660, September 1960.

[3] R. L. Kyle, "Plotting impedances with negative resistive components," *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-8, p. 377, May 1960.

[4] L. J. Kaplan and D. J. R. Stock, "Some comments on the method of Kyhl," *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-8, p. 668, November 1960.

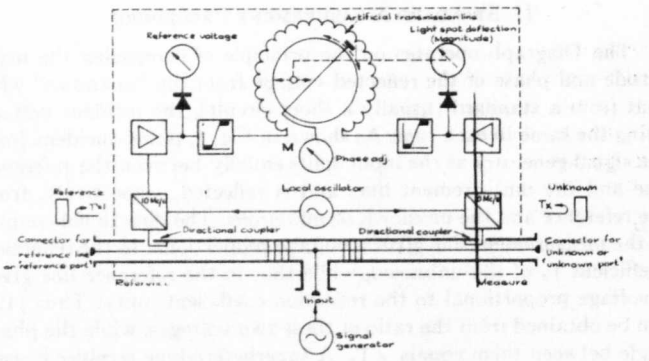


Fig. 1. Schematic of Z-g Diagraph.

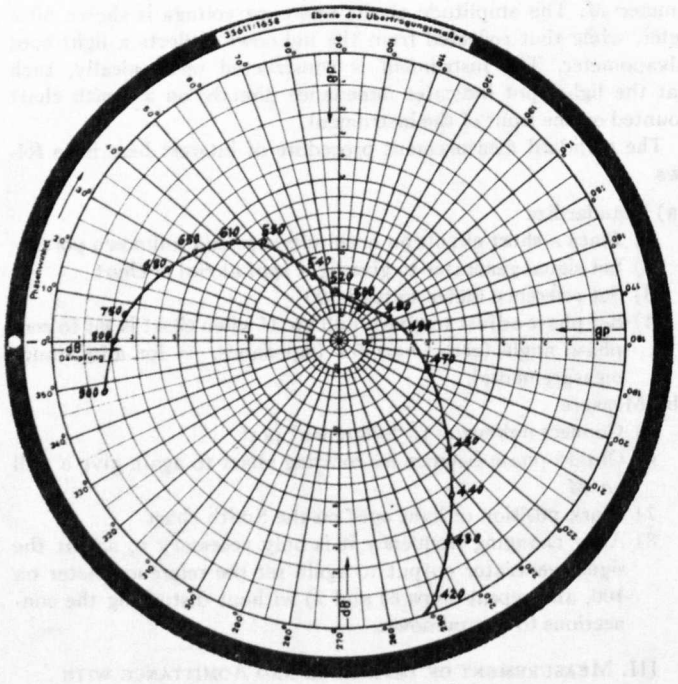


Fig. 2. Gain and phase vs. frequency, for a tunnel diode reflection amplifier.

7) Obtain phase balance by rotating chart to again give a null on M
8) Mark position of light spot on the Smith chart
9) After changing frequency adjust signal generator output to again set the reference meter on 100, and repeat steps 7) and 8).

Reflection gain is measured by simply changing the Smith chart on the front of the instrument, for another type of chart. Supplied for attenuation measurements on two-ports is a polar chart[1] calibrated in dB and phase angle. It so happens that this chart, used with the measurement procedure 1)–9) above, conveniently indicates reflection gain G of the one-port active device, where

$$|G| = 20 \log_{10} \left| \frac{Z_x + R_0}{Z_x - R_0} \right| \text{db}$$

$$\underline{/G} = -\underline{/\Gamma_x}$$

and R_0 is the characteristic resistance of the Diagraph transmission lines = 50 ohms. Z_x is the impedance of the unknown. Such a plot of reflection gain for a tunnel diode amplifier is shown in Fig. 2.

IV. Discussion

When measuring active one-ports precautions are required to avoid instability and overloading of the item being tested.

The reflection coefficient seen looking in at either the reference or the unknown ports of the Diagraph, referred to 50 ohms, is a maximum of 0.005 for frequencies 30 Mc/s to 420 Mc/s and 0.01 for frequencies up to 4.2 Gc/s. Thus if the unknown is stable when connected to a good 50 ohm matched load it will be stable when connected to the Diagraph. This obviates one of the difficulties experienced when using a slotted line to measure negative impedances.

The instrument is generally sufficiently sensitive to avoid overloading low power tunnel diode amplifiers, though some overloading has been experienced when attempting to operate tunnel diodes close to their cutoff frequency. It is possible to increase the sensitivity of the Diagraph by a factor of 3 in such cases, with some loss of accuracy in the amplitude measurement, but this can then be regained by using a signal generator with a calibrated attenuator as a signal source. No trouble from overloading should be experienced with parametric amplifiers.

M. K. McPhun
Mullard Research Labs.
Salfords, Surrey
England

APPENDIX 7.

U.H.F. tunnel-diode amplifier.

Proceedings I.E.E. 114, 4, pp428-434, (April 1967).

U.H.F. tunnel-diode amplifier

M. K. McPhun, B.Sc., C.Eng., M.I.E.E.

Synopsis

The paper is mainly devoted to a detailed account of the design of the reflection amplifiers for use in a tunnel-diode repeater described in a companion paper. The amplifiers are constructed in stripline, incorporating transmission lines of variable characteristic impedance. Thus tunnel diodes of widely differing characteristics can be accommodated in a single amplifier; yet the tuning procedure is particularly simple. Results are also given for a circulator-coupled amplifier formed from the reflection amplifier and a 4-port circulator.

1 Introduction

The reflection amplifier described here was developed for use in the two types of radio-repeater amplifiers described in a companion paper.[1] To achieve simultaneous amplification in two directions at two different frequencies, the repeater amplifiers use 4-port circulators, filters and two reflection amplifiers tuned to 454·2 and 468·2 Mc/s, respectively. At the time this work was undertaken, most work on tunnel-diode amplifiers was of a trial-and-error nature; moreover, tunnel diodes were selected to give the required gain in an amplifier.[2] A main object of the work was therefore that the amplifiers should be designable; it would follow that the designs could be reproduced at higher frequencies if required at a later date. This paper is mainly about the design of the reflection amplifiers, but, as they are of application to many other purposes apart from the repeater, some typical results on a circulator-coupled amplifier using one of these reflection amplifiers are included.

The requirements for the reflection amplifiers are as follows:

(*a*) Only a narrow bandwidth, of the order of 1 Mc/s, is needed, but a large bandwidth is acceptable as the repeater bandwidth is limited by the directional or the diplexing filters.
(*b*) The amplifiers must have negligible gain at 3 times the centre frequency, where the second resonance of the directional filters occurs.
(*c*) The centre frequency must be variable by about ±10 Mc/s either side of 461 Mc/s.
(*d*) The amplifier gain must be adjustable between about 15 and 30 dB without changing the tunnel-diode bias.
(*e*) The amplifiers must accept a range of tunnel diodes without requiring close selection of some of their parameters.
(*f*) It is, of course, desirable that the amplifiers should be easy to tune, e.g. if the tunnel diodes are changed.

Before describing the amplifier, it is desirable to summarise some background theory on the tunnel diode.

2 Theory

The tunnel diode has a voltage-controlled negative-resistance characteristic whose main parameter is the peak current. All the work described here was done using tunnel diodes with a peak current of about 2 mA, as these are most suitable for use in conjunction with 50Ω-transmission-line circuits. When suitably biased in the negative-resistance region, the tunnel diode has the equivalent circuit of Fig. 1*a*(i).[3] The junction of the tunnel diode is represented by a negative resistance $-R_d$ (or conductance $-G_d$) in parallel with the junction capacitance C_d, where R_d is a positive number. r and L represent the unavoidable resistance and inductance in series with the diode. R_d varies with the bias point throughout the negative-resistance region, being a minimum at the inflection point. Biasing the tunnel diode at the inflection point will give an amplifier which is least sensitive to variations in the bias voltage, and throughout it is assumed that the tunnel diode is biased at this point unless otherwise stated.

2.1 Frequency response of tunnel diode

The input impedance $Z(\omega)$ in Fig. 1 *a*(i) is a function of frequency, given by

$$Z(\omega) = r - \frac{G_d}{G_d^2 + \omega^2 C_d^2} + j\left(\omega L - \frac{\omega C_d}{G_d^2 + \omega^2 C_d^2}\right) \quad (1)$$

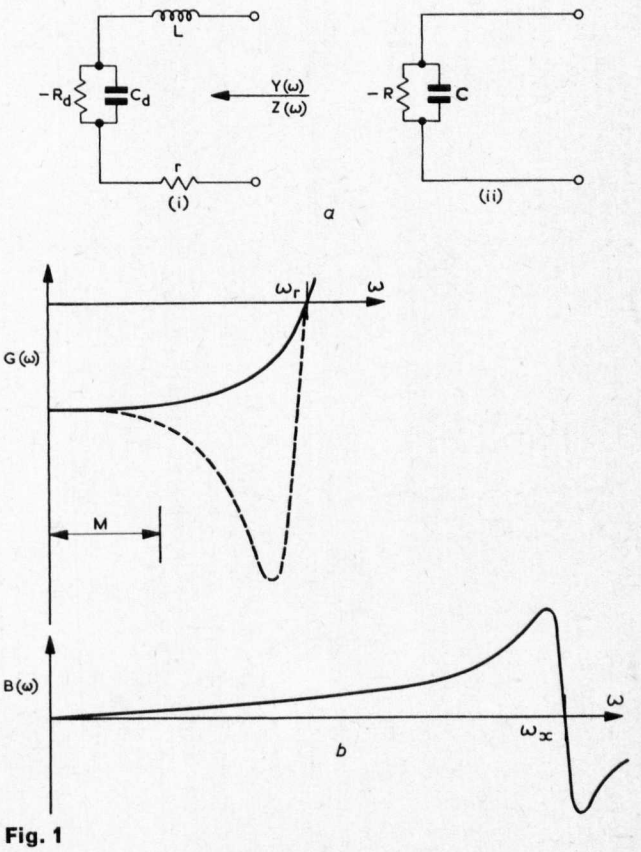


Fig. 1

Performance of tunnel diode

a Equivalent circuits of the diode
(i) Equivalent circuit
(ii) Approximate equivalent circuit, valid in the region M
b Frequency response of $G(\omega)$ and $B(\omega)$
M extends to about $\omega_r/3$

It is usually more convenient to consider the conductance $Y(\omega)$, where

$$Y(\omega) = G(\omega) + jB(\omega) = \frac{1}{Z(\omega)} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The frequency response of $G(\omega)$ and $B(\omega)$ is shown in Fig. 1*b*, where the resistive cutoff frequency ω_r and the series resonant frequency ω_x are illustrated.[4, 13] For $\omega_x > \omega_r$ the condition

$$L < C_d R_d r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

is obtained. A tunnel diode fulfilling this condition is known as 'absolutely stable';[4] Fig. 1 has been drawn for ω_x about twice ω_r. For this condition, and for higher ω_x, $G(\omega)$ will follow the full line in Fig. 1. However, if ω_x is reduced to approach ω_r, for instance by increasing L, a peak will appear in the $G(\omega)$ curve, as shown dotted in Fig. 1*b*. For ω_x in the region of ω_r, this peak may exceed G_d by many times, and this condition makes the design of stable amplifiers

Paper 5217 E, first received 18th August and in revised form 25th November 1966
Mr. McPhun was formerly with the Central Electricity Research Laboratories, Leatherhead, Surrey, England, and is now with the School of Engineering Science, University of Warwick, Coventry, War., England

extremely difficult. Throughout the following work, tunnel diodes fulfilling the condition $\omega_x \geqslant 2\omega_r$ were used, so that the behaviour of $G(\omega)$ was given by the full line in Fig. 1*b*. Great care had to be taken when using diodes not to increase their inductance L by the way they were mounted in the circuit.

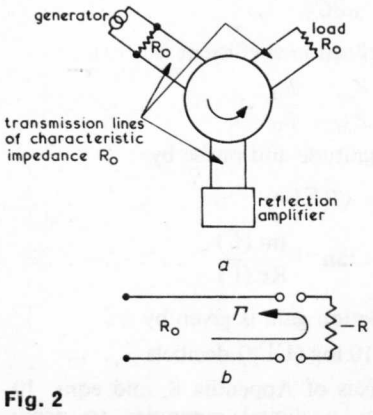


Fig. 2

Details of reflection amplifier

a Basic circulator-coupled amplifier
b Reflection amplifier

Over the frequency range M in Fig. 1*b*, from zero to about $\omega_r/3$, the equivalent circuit of the tunnel diode may be approximated, as shown, to a negative resistance $-R$ shunted by a capacitance C. For the tunnel diodes of interest here, R and C differ little from R_d and C_d when $\omega < \omega_r/3$. It is, however, essential to use the full equivalent circuit and eqn. 2 when considering the performance of circuits at higher frequencies than $\omega_r/3$, e.g. when examining stability.

When wide bandwidth is not a primary consideration, too high an ω_r is a disadvantage. Such a high resistive cut-off frequency merely increases the range of frequency over which the amplifier must be prevented from oscillating. A useful rule of thumb for most amplifiers is to make ω_r about 3 times the design frequency of the amplifier. This places the design frequency at the upper end of region M in Fig. 1*b*, so that the second resonance of associated transmission-line resonators falls above ω_r. Unfortunately, we were not able to obtain tunnel diodes that were suitable in this respect for these amplifiers.

2.2 Tunnel diode used in a reflection amplifier

The basic form of a circulator-coupled tunnel-diode amplifier is as shown in Fig. 2*a*, where the circulator isolates the reflection amplifier from both the generator and the load. The reflection amplifier (Fig. 2*b*) terminates the transmission line of characteristic resistance R_0 in a negative resistance $-R$, giving a reflection coefficient Γ which is greater than unity.

$$\Gamma = \frac{(-R) - R_0}{-R + R_0} = \frac{R + R_0}{R - R_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

It is convenient to use the reflection power gain A, expressed in dB:

$$A = 10 \log |\Gamma|^2 \text{dB} = 20 \log \frac{R + R_0}{R - R_0} \quad . \quad . \quad . \quad (5)$$

which is also the transmission gain of the circulator-coupled amplifier of Fig. 2*a*.

To use a tunnel diode as a reflection amplifier we must do the following four things:

(*a*) Bias the tunnel diode, using a load line providing a single-valued intersection in the negative-resistance region of the I/V characteristic
(*b*) Present the transmission line with a suitable negative resistance at the centre frequency f_0
(*c*) Reduce Γ to acceptable limits at frequencies away from f_0
(*d*) Maintain stability while achieving (*a*), (*b*) and (*c*).

The last requirement of stability is of overriding importance, and is likely to occupy much of the effort of the designer.

2.3 Stability

The tunnel diode presents a negative resistance at its terminals from d.c. to the resistive cut-off frequency ω_r. Therefore it is necessary to consider the behaviour of a tunnel-diode circuit throughout this frequency range, and not only in the frequency range of interest. The components used in the circuit must not have spurious resonances below ω_r, or, if they have, the effect of these resonances must be considered.

Much of the difficulty in ensuring stable operation is caused by the finite bandwidths of the circulator, generator and load. For the first repeater amplifier, these problems were solved by the use of the directional filters[1] interposed between the reflection amplifier and the circulator in Fig. 2, resulting in the reflection amplifier 'seeing' the resistance R_0 from d.c. to $3f_0$. Thus the stability requirements for the reflection amplifier were reduced to ensuring that the amplifiers were stable when connected to a nominal 50Ω-characteristic-resistance transmission line over the frequency range 0 to $3f_0$. These requirements were in fact more severe than necessary, and, in the final repeater configuration, stable operation was obtained with the reflection amplifier connected directly to a circulator. The author has described a stability prediction technique[5] that requires only plots of $|\Gamma|$ against frequency. This technique was employed to examine any reflection amplifier for possible oscillations before it was made, and obtain an indication of the cause of the instability.

3 Reflection amplifier

A schematic diagram of the amplifier is shown in Fig. 3*a*. It is based on an amplifier described by Reindel,[4] and was chosen for its comparatively narrow bandwidth and inherent stabilising circuit.

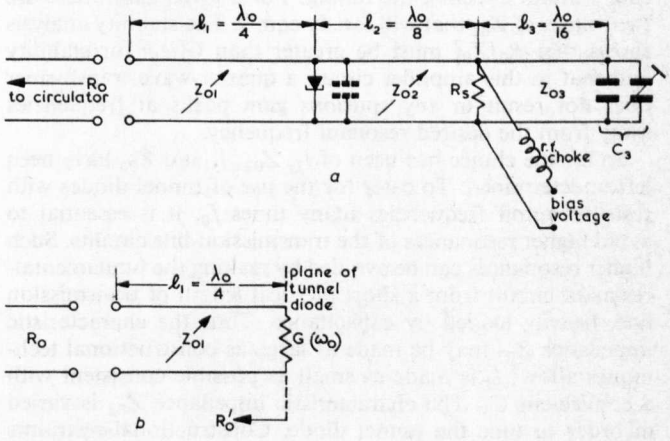


Fig. 3

Circuits of reflection amplifier

a Schematic diagram of reflection amplifier
b Effective equivalent circuit at resonance

3.1 Amplifier theory

Consider the circuit to the right of the tunnel diode in Fig. 3*a*. The tunnel diode is placed across a transmission line of characteristic impedance Z_{02}. A resistance R_s shunts the transmission line one eighth of an electrical wavelength from the tunnel diode (at the design frequency ω_0); this resistance serves the dual purpose of providing a d.c. load for the tunnel-diode-bias supply, and acting as a stabilising resistance. A further length of transmission line of characteristic impedance Z_{03}, connected in parallel with R_s, is terminated in a capacitor C_3.

The principle of operation is as follows. At the design frequency ω_0, the transmission line Z_{03} loaded with the capacitor C_3 is resonant, and presents a short circuit across the resistance R_s. Thus transmission line Z_{02} can be made to resonate with C, the effective capacitance of the tunnel diode. Therefore, at the resonant frequency ω_0, the amplifier appears simply as $G(\omega_0)$ terminating the transmission line Z_{01}, where $G(\omega_0)$ is given by eqn. 2. At all frequencies away from resonance, R_s provides damping to the circuit as it no longer appears short-circuited. Analysis of the stability of this circuit indicated a noncritical optimum R_s of $R_d/3$. The length l_3 is given by

$$l_3 = \frac{c}{360 f_0} \tan^{-1}\left(\frac{1}{2\pi f_0 C_3 Z_{03}}\right) \text{metres} \quad . \quad . \quad . \quad (6)$$

where c is the velocity of light (3×10^8 m/s), $\tan^{-1}\left(\frac{1}{2\pi f_0 C_3 Z_{03}}\right)$ is expressed in degrees, $f_0 = \omega_0/2\pi$ and C_3 and Z_{03} are in farads and ohms, respectively. The length l_2 is given by

$$l_2 = \frac{c}{360 f_0} \tan^{-1}\left\{\frac{1}{B(\omega_0)Z_{02}}\right\} \text{ metres} \quad . \quad . \quad . \quad . \quad (7)$$

where $B(\omega_0)$ is found from eqn. 2. This will ensure that at the resonant frequency f_0 the admittance seen at the terminals of the tunnel diode is real and equal to $G(\omega_0)$, as given by eqn. 2. At the resonant frequency, the amplifier appears as shown in Fig. 3*b*. l_1 is a quarter wavelength at this frequency, forming a quarter-wave transformer and transforming the characteristic impedance R_0 to some value R_0' across the tunnel diode, where

$$R_0' = \frac{Z_{01}^2}{R_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

From eqns. 8 and 5 we obtain the gain A of the amplifier at resonance as follows:

$$A = 20 \log \frac{\frac{R_0}{Z_{01}^2} - G(\omega_0)}{\frac{R_0}{Z_{01}^2} + G(\omega_0)} \text{ decibels} \quad . \quad . \quad . \quad . \quad (9)$$

where $G(\omega_0)$ is a negative number. By making the characteristic impedance of the quarter-wave transformer Z_{01} variable, the gain of the amplifier may be controlled with ease without affecting the tuning. For a given gain, there are two values of Z_{01} that will satisfy eqn. 9. The stability analysis shows that R_0/Z_{01}^2 must be greater than $G(\omega_0)$ for stability and that in this amplifier circuit a quarter-wave transformer does not result in any spurious gain peaks at frequencies away from the desired resonant frequency.

So far the choice has been of l_2, Z_{02}; l_3 and Z_{03} have been left undetermined. To cater for the use of tunnel diodes with resistive cutoff frequencies many times f_0, it is essential to avoid higher resonances of the transmission-line circuits. Such higher resonances can be avoided by making the fundamental-resonant circuit from a short physical length of transmission line, heavily loaded by capacitance. Thus the characteristic impedance Z_{03} may be made as large as constructional techniques allow; l_3 is made as small as possible consistent with a convenient C_3. The characteristic impedance Z_{02} is varied in order to tune the tunnel diode. Constructional restraints on the range of Z_{02} require l_2 to be about one eighth of a wavelength long.

3.2 Amplifier frequency response

The formula for the gain of the amplifier as a function of frequency is very complicated, so that all calculations were done in terms of a product of chain matrixes for the individual sections of the amplifier, as shown in Fig. 4. The

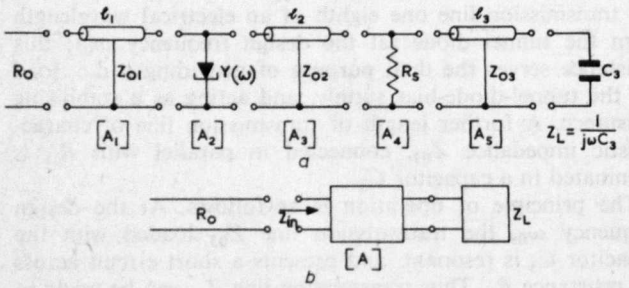


Fig. 4

Matrixes for calculation of amplifier frequency response

a Individual matrixes
b Equivalent matrix

product of the five matrixes is represented by an equivalent matrix A, as in Fig. 4*b*, where

$$A = [A_1][A_2][A_3][A_4][A_5] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (10)$$

The matrixes A_1 to A_5 are listed in Appendix 8.

The amplifier input impedance is then given by

$$Z_{in} = \frac{A_{11}Z_L + A_{12}}{A_{21}Z_L + A_{22}} = \text{Re}\,(Z_{in}) + j\,\text{Im}\,(Z_{in}) \quad . \quad (11)$$

where all the quantities are complex, and

$$Z_L = \frac{1}{j\omega C_3}$$

The input reflection coefficient is

$$\Gamma = \frac{Z_{in} - R_0}{Z_{in} + R_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

given in magnitude and phase by

$$|\Gamma| = \sqrt{(\Gamma\Gamma^*)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$\angle\Gamma = \tan^{-1}\frac{\text{Im}\,(\Gamma)}{\text{Re}\,(\Gamma)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and the reflection gain is given by

$$A = 10 \log (\Gamma\Gamma^*) \text{ decibels} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The matrixes of Appendix 8, and eqns. 10–15, were programmed for a digital computer to produce frequency-response curves.

3.3 Construction of amplifier

460 Mc/s lies within that awkward frequency band where lumped circuits become uncomfortably small, while transmission-line circuits tend to be long and unwieldy. The first choice lay between lumped and transmission-line circuits; the latter were chosen to allow for scaling of the design to higher frequencies, if required at some future date. Then it was necessary to choose the type of transmission line. A first design was constructed in coaxial line, with fixed characteristic impedances and fixed tuning; this design worked satisfactorily, but was difficult to construct and alter. It will not be discussed further.

For the subsequent work high-Q triplate strip transmission line,[6] henceforth referred to as stripline, was chosen for the following reasons:

(*a*) The bias supply and resistor R_s are much easier to arrange than in the coaxial construction.
(*b*) Junctions between transmission lines are easier to construct in stripline.
(*c*) The following ways of making transmission lines with a variable characteristic impedance could be envisaged.

The construction of the variable-impedance striplines is shown in Fig. 5. Fig. 5*a* shows the conventional stripline:

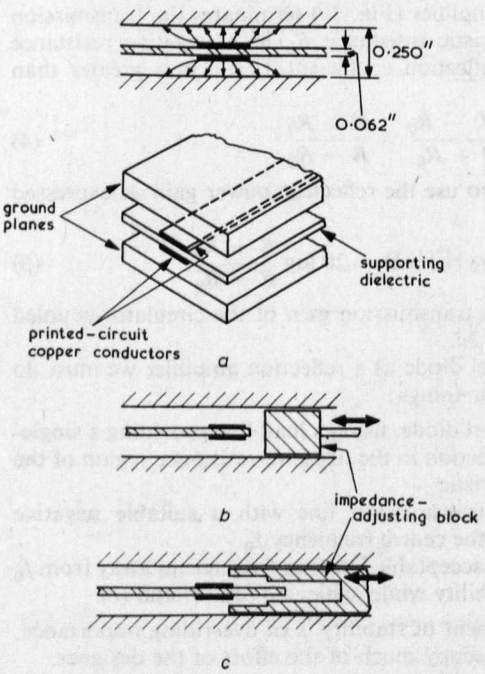


Fig. 5

Construction of variable-impedance striplines

a Conventional high Q triplate stripline
b and *c* Two methods of varying Z_0

a TEM transmission line in which the inner conductor is symmetrically placed between two ground planes. The inner conductor takes the form of a printed circuit of identical shape on both sides of the dielectric. All current flow in both the ground planes and the inner conductors is longitudinal, and only a small fringing field is within the dielectric. In Fig. 5*b*, the characteristic impedance is decreased by inserting a conducting strip between the ground planes to one side of the inner, increasing the capacitance per unit length. Movement of the strip towards the inner results in a decrease in the characteristic impedance. There is no problem with contact between the block and the ground planes as all current flow is longitudinal. Fig. 5*c* shows an alternative way of varying the characteristic impedance by effectively altering the spacing between the inner and the two ground planes. Again, this reduces the characteristic impedance as the block is moved towards the inner conductor, by increasing the capacitance per unit length. By suitably shaping this block, a wide variety of curves of characteristic impedance against block position can be obtained. The final version of the reflection amplifier is shown in Fig. 6 with one of the ground

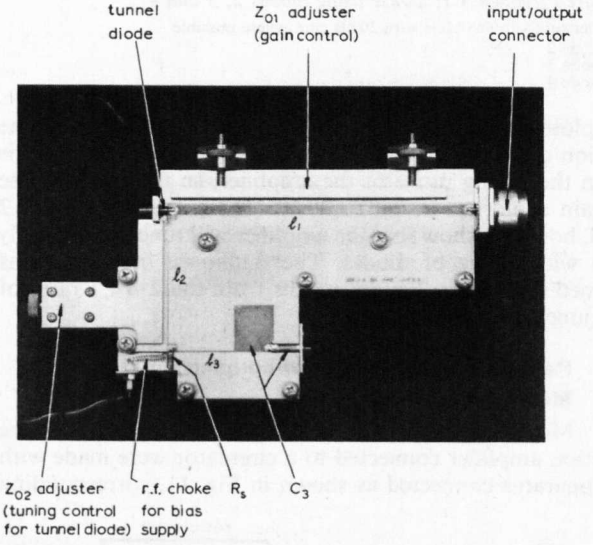


Fig. 6

Reflection amplifier with one ground plane removed

planes removed. It was constructed with two right-angle bends in the strip transmission line in order to facilitate mounting the lumped components, and to make the amplifier compact. Experience had shown that it was essential to construct the whole amplifier symmetrically about the inner conductor, as any asymmetry could lead to the formation of parallel-plate modes between the two ground planes, which could cause oscillations.

The variable quarter-wave-transformer gain control uses the impedance adjuster of Fig. 5*b*; this enables Z_{01} to be adjusted over the range 50–35Ω. The pill-type tunnel diode is mounted on end to the centre conductor to preserve symmetry. To tune the diode, the impedance adjuster of Fig. 5*c* is used to vary Z_{02}. The widest range of Z_{02} that could conveniently be obtained with this construction was from about 30 to 100Ω. Insertion of these values into eqn. 7 showed that a length $l_2 = 7{\cdot}4$ cm enabled tunnel diodes with capacitances in the range 3–9 pF to be tuned. The bias resistor R_s consists of a metal-oxide resistor mounted symmetrically between the ground planes. Measurements on such resistors showed that their inductance was negligible up to a frequency of 1000 Mc/s, when mounted in this way. Z_{03} was made 100Ω, this being the highest value obtained conveniently with a strip width of 0·5 mm. Length l_3 was made as short as possible by using the highest C_3 permitted by this method of construction. C_3 is in two parts: a fixed part of 5 pF consisting of a short length of 18Ω stripline and a variable trimmer adjustable from 1 to 7 pF. Lengths l_1, l_2, l_3 had to be made shorter than calculated for free space, to allow for the lower velocity of propagation in the stripline.[6]

4 Measurements, results and discussion

Much effort was expended upon the reflection amplifier prior to its incorporation into the repeater, and the results shown in Section 4.1 for the reflection amplifier are a small selection from those available. On the other hand, the experiments described in Section 4.2, using the reflection amplifier with a 5% bandwidth circulator, are included as an example of what can be achieved with the reflection amplifier. They should not be regarded as optimum results, as no time was available for improving the characteristics obtained at the first attempt.

4.1 Performance of reflection amplifier

Tuning these amplifiers is particularly simple, the procedure being as follows. With the tunnel diode removed, the reflection coefficient of the reflection amplifier is monitored at frequency f_0, and C_3 is adjusted to give unity reflection coefficient. Then, with the diode inserted, Z_{02} is adjusted to give maximum gain at f_0, and finally the gain is adjusted to the desired value using Z_{01}. When tuning is carried out in this order, the controls are completely independent of each other.

4.1.1 Measurement methods

All measurements on the reflection amplifiers were made using Z–g diagraphs[7] having frequency ranges from 30 Mc/s to 2400 Mc/s. The author has shown[8] how these instruments may be used for measuring reflection coefficients greater than unity by interchanging the normal 'reference' and 'measure' terminals. This method terminates the amplifier in a good matched load, and operates at a sufficiently low power level so as to not saturate the amplifier. The results obtained are the magnitude and phase of the reflection coefficient plotted on the negative Smith chart,[9] which is the transformation $\Gamma' = 1/\Gamma$ of the normal Smith chart. Thus the reflection coefficient rises to infinity at the centre of the chart, where $R = R_0$ (see eqn. 4), and is unity at the rim. Alternatively, the results could be plotted directly on a chart with the radial scale calibrated to read reflection gain in decibels. The facility of being able to plot phase was invaluable. Even where there is a large narrow gain peak, this type of presentation gives a smooth curve; the presence of a large gain is merely indicated by the curve passing near the centre of the chart. Any extraneous loops in the curve invariably indicated maloperation of the amplifier.

The stability of the amplifier was checked after every alteration to it, however slight. This was done very simply, using the bias arrangement in Fig. 7 whereby the d.c. supply

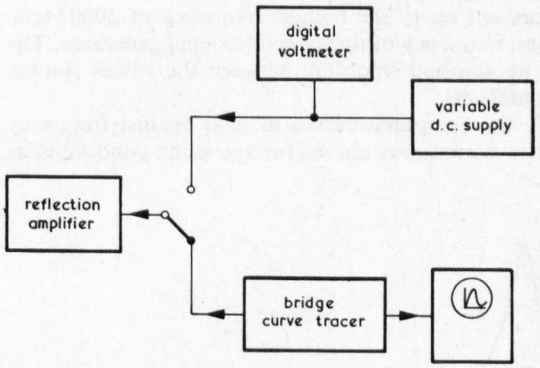


Fig. 7

Bias arrangement for tests on reflection amplifiers

to the tunnel diode could be interchanged with an a.c. sweep from a bridge-type curve tracer.[10] By switching to the curve tracer, the current/voltage characteristic of the tunnel diode could be displayed with the diode operating in the amplifier, and any oscillation immediately observed as a distortion of this characteristic.[11] The characteristic was also a sensitive indicator of the presence of overloading due to too large a signal amplitude.

4.1.2 Results

For convenience, the characteristics of the tunnel diodes used in these experiments are summarised in Table 1. Fig. 8 shows the reflection gain in dB and phase angle for an amplifier tuned to 454 Mc/s using tunnel diode 1. Curves are plotted for two positions of the gain control, no tuning

Table 1

CHARACTERISTICS OF TUNNEL DIODES USED IN EXPERIMENTS

Diode	I_p	R_d	C_d	r	f_r
	mA	Ω	pF	Ω	Mc/s
1	2·1	58	6·4	1·5	2640
2	1·85	58·4	7·0	0·4	4770
3	2·1	57·2	2·9	1·2	6570
4	1·5	68·5	2·3	2·5	5300

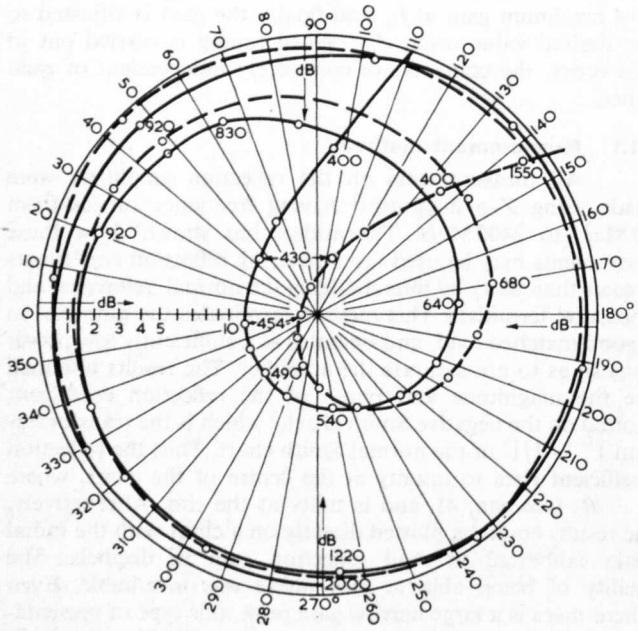


Fig. 8

Gain and phase of reflection amplifier; gain control using λ/4 transformer

—— $Z_0 = 50\,\Omega$ (no transformer)
---- $Z_0 < 50\,\Omega$
Figures along curves are frequencies in Mc/s

adjustments being made between plots. No auxiliary gain peaks are present up to the highest frequency of 2000 Mc/s, the maximum frequency of the available signal generator. The gain could be adjusted smoothly between the values shown, i.e. 12dB and 24dB.

In Fig. 9, the computed curves of gain against frequency are plotted as continuous curves for the same conditions as

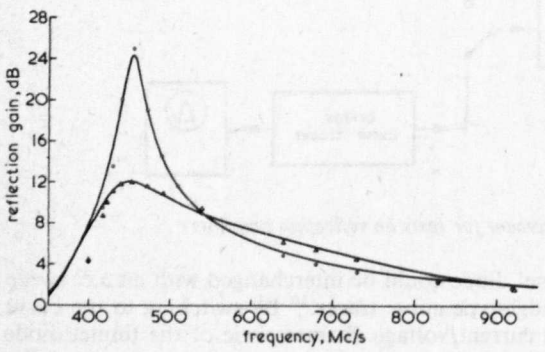


Fig. 9

Computed frequency response for two positions of gain control

○ △ experimental points taken from Fig. 8

Fig. 8. Experimental points from Fig. 8 are also plotted in Fig. 9 for comparison with the theory. It is seen that agreement is very close, with the exception of points below resonance. This discrepancy was due to the presence of the r.f. inductor and the bias-supply leads (Figs. 3*a* and 6), which were not included in the theoretical model. It was considered adequate that the r.f. inductor had sufficient inductance not to affect the gain in the region of f_0. Too many turns on the r.f. inductor would have resulted in spurious selfresonance at higher frequencies with consequent danger of instability.

The effect upon the amplifier gain/frequency response of changing diodes is shown in Fig. 10. The frequency response

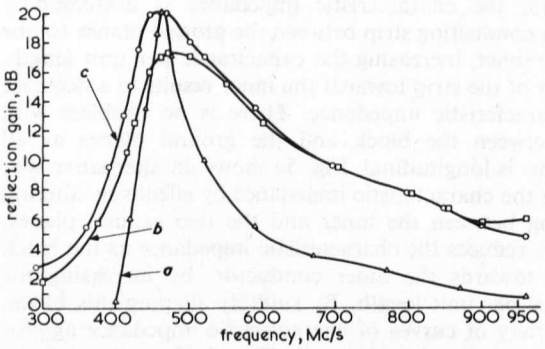


Fig. 10

Measured frequency response using diodes 2, 3 and 4

Amplifier tuned to 468 Mc/s with 20 dB gain where possible
a Diode 2
b Diode 3
c Diode 4

was plotted with diodes 2, 3 and 4 in the amplifier. The junction capacitance of diode 4 (2·3pF) was too small to lie within the tuning range of the amplifier, so that in this case the gain could not be set to 20dB. The curves for diodes 2 and 3, however, show that the amplifier will tune satisfactorily for a wide range of diodes. The difference in bandwidths obtained for diodes 2 and 3 results from the 2·4 : 1 ratio of their junction capacitances.

4.2 Performance of circulator-coupled amplifier

4.2.1 Measurement methods

Measurements of the gain/frequency response of the reflection amplifier connected to a circulator were made with the apparatus connected as shown in Fig. 11, corresponding

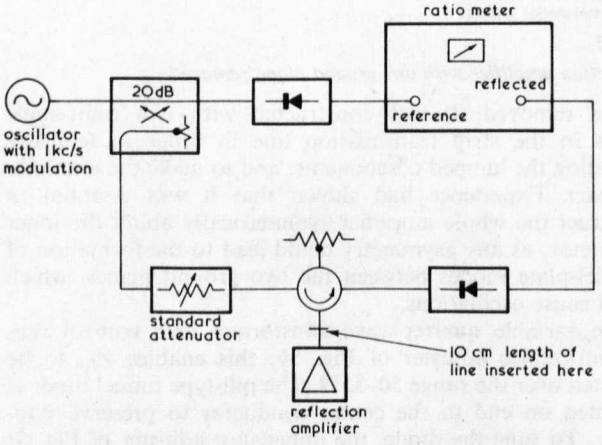


Fig. 11

Measurement of frequency response of circulator-coupled amplifier

to the type II repeater,[1] which enables the signal applied to the tunnel diodes to be kept sufficiently low to avoid overloading. The ratio of output to reference-signal levels, as indicated on the ratio meter, was kept constant by adjustment of the standard attenuator, which indicated the amplifier gain. The noise figure of the amplifier was measured using a diode-noise generator as a calibrated noise source. For a receiver a field-intensity meter was used, with its d.c. output monitored with a digital voltmeter. The receiver and the digital voltmeter were calibrated using a c.w. signal; this procedure was considered to be justified in view of the quoted peak-handling properties of the detector in the field-intensity meter.

4.2.2 Results

The reflection coefficient Γ_c of one port of the circulator, with the remaining three ports terminated in matched loads, is shown in Fig. 12. This was measured using the diagraph.

The circulator has a nominal bandwidth of 5%, but in fact the bandwidth between the $|\Gamma_c| = 0{\cdot}1$ points is 14%. The curve of Fig. 12 is rather unusual in that it is more common for the loop to enclose the origin; thus this circulator has a considerable reactance at its centre frequency.

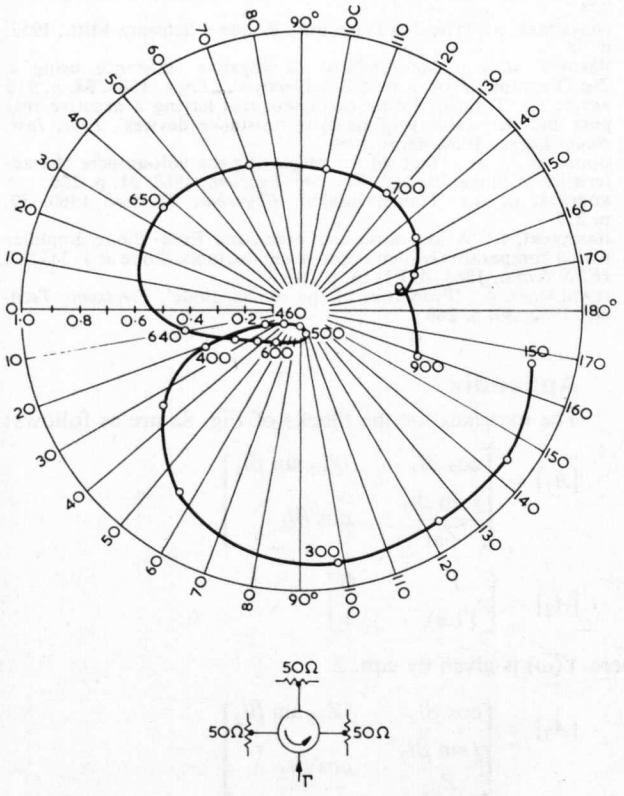


Fig. 12

Reflection of circulator terminated at three ports

Figures along curve are frequencies in Mc/s

The frequency response of Fig. 13*a* was obtained with the amplifier connected directly to the circulator. The plateau at 6dB gain results from the out-of-band performance of the circulator. This plateau could be eliminated by inserting a 10cm length of 50Ω coaxial line between the circulator and amplifier, giving the more acceptable response of Fig. 13*b*. Hamasaki[12] has shown how the resonance of the circulator can be used for reactance compensation to broaden the bandwidth of the reflection amplifier. This could not be done simply by correctly spacing the amplifier from the circulator in this case, because of the asymmetry of the circulator characteristic. For curve *b* in Fig. 13, the 3dB bandwidth B is 5·8%, giving a gain–bandwidth product[13] $(\sqrt{G} - 1)B =$ 260 Mc/s. The corresponding figure for the same tunnel diode tuned by a single lumped inductance would be[13] $1/(\pi R_d C_d) =$ 780 Mc/s. Thus this design has restricted the bandwidth considerably. The amplifier was stable for all source and load mismatches.

Graphs of the measured total system noise figure, measured

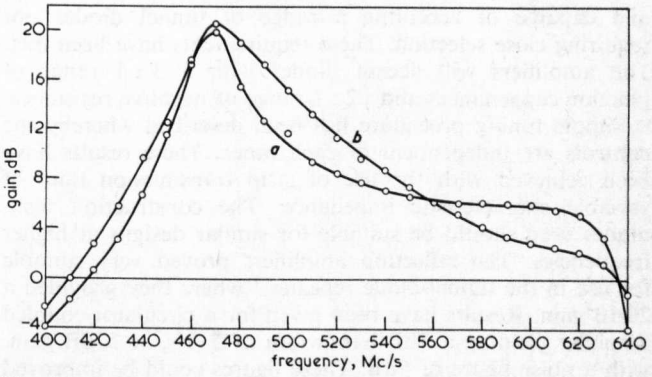


Fig. 13

Gain/frequency curve for circulator-coupled amplifier

Diode 2 Bias: 97 mV

a Amplifier directly connected to circulator

b 10 cm length of transmission line between amplifier and circulator

amplifier gain, and the amplifier noise figure calculated from these results, are shown in Fig. 14 as a function of tunnel-diode bias, for diodes 1 and 2. The vertical lines on the graph of total system-noise figure are an estimate of the probable experimental error. The minimum amplifier noise figure is obtained at a higher bias voltage than that giving maximum gain, which agrees with theory.[12, 13] The noise figure N for the tunnel diode may be calculated[13] from

$$N = \frac{1 + 20I_0R_d}{\left(1 - \frac{r}{R_d}\right)\left\{1 - \left(\frac{f_0}{f_r}\right)^2\right\}}$$

where I_0 is the bias current. To this must be added the 0·3dB insertion loss of the circulator. Table 2 compares the calculated and measured figures.

Table 2

MINIMUM NOISE FIGURE

	Calculated	Measured
	dB	dB
Diode 1 .	4·47	5·3
Diode 2 .	3·9	5

This noise performance was adequate for the repeater application, and no time was available for attempts to improve the noise figure or to investigate the reason for this discrepancy.

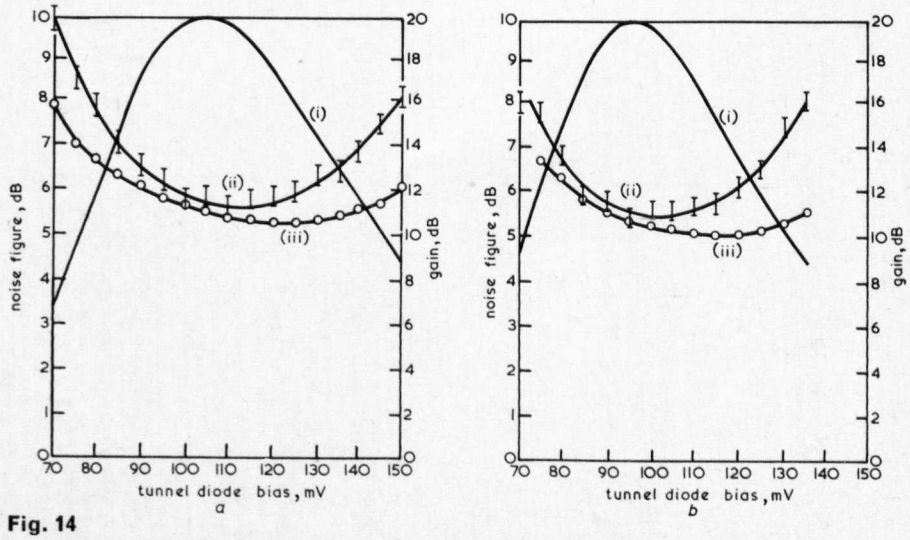


Fig. 14

Noise figures and gain plotted against tunnel-diode bias

a Diode 1

b Diode 2

(i) Tunnel-diode amplifier gain

(ii) Total system noise figure

(iii) Tunnel-diode amplifier noise figure

The saturation characteristics of the amplifier are discussed in full in the companion paper.[1]

5 Conclusions

Throughout this work a main aim has been that the tunnel-diode amplifiers should be designable, easy to set up, and capable of accepting a range of tunnel diodes not requiring close selection. These requirements have been met. The amplifiers will accept diodes with a 3 : 1 range of junction capacitances and a 2 : 1 range of negative resistance. A simple tuning procedure has been described whereby the controls are independent of each other. These results have been achieved with the use of strip transmission lines of variable characteristic impedance. The construction techniques used should be suitable for similar designs at higher frequencies. The reflection amplifiers proved very suitable for use in the tunnel-diode repeater,[1] where they provided a 20dB gain. Results have been given for a circulator-coupled amplifier yielding a 3dB bandwidth of 5·8% at 20dB gain, with a noise figure of 5dB. These figures could be improved by using a more suitable circulator; that used was one of the first u.h.f. 4-port circulators available. Amplifiers of this type would be easier to design and make at higher frequencies, e.g. above 1500Mc/s, as the tunnel diodes available are intended for such microwave operation.

6 Acknowledgments

The author wishes to acknowledge the assistance of G. R. Frances and H. E. Spencer in carrying out measurements, and that of G. N. Porter and G. Berg in the construction of the reflection amplifiers. Thanks are due to the CEGB computing branch for programming effort, to J. H. Longley of Cossor Electronics Ltd. for help with the stripline techniques, and to M. A. Lee of Associated Semiconductors Manufacturers Ltd. for the loan of experimental tunnel diodes. Thanks are also due to Mullard Research Laboratories for preparing the drawings and typing the manuscript. This work was performed at the Central Electricity Research Laboratories, and is published by permission of the Central Electricity Generating Board.

7 References

1 HEDDERLY, D., HOOPER, J., and McPHUN, M. K.: 'Short-hop radio-relay systems using tunnel-diode repeaters.' (see next page)
2 REINDEL, J.: 'A compact tunable tunnel diode S-band receiver', *Microwave J.*, 1961, **4**, p. 92
3 HALL, R. N.: 'Tunnel diodes', *IRE Trans.*, 1960, **ED–7**, pp. 1–9
4 REINDEL, J.: 'Tunnel-diode circuits at microwave frequencies', US Government Research Report, ASTIA No. AD 269846, 1961
5 McPHUN, M. K.: 'Practical stability criterion for tunnel diode circuits', *Electronics Letters*, 1965, **1**, pp. 167–168 and p. 289
6 FOSTER, K.: 'The characteristic impedance and phase velocity of high-Q triplate line', *J. Brit. Instn. Radio Engrs.*, 1958, **18**, pp. 715–723
7 EICHACKER, R.: 'The Z–g Diagraph', Rohde u Schwarz Mitt., 1952, p. 78
8 McPHUN, M. K.: 'Measurement of negative resistance using a Z-g Diagraph', *Proc. Inst. Elect. Electronics Engs.*, 1966, **54**, p. 910
9 ROSIN, B.: 'Transformation of impedances having a negative real part and the stability of negative-resistance devices', *Proc. Inst. Radio Engrs.*, 1960, **48**, p. 1660
10 GOODMAN, A. M.: 'Test set for displaying the volt-ampere characteristics of tunnel diodes', *Rev. Sci. Instrum.*, 1960, **31**, p. 286
11 ROBERTS, G. N.: 'Tunnel diodes', *Electronic Technol.* 1960, **37**, p. 217
12 HAMASAKI, J.: 'A low-noise and wide-band Esaki diode amplifier with a comparatively high negative conductance diode at 1·3 Gc/s', *IEEE Trans.*, 1965, **MTT–13**, p. 213
13 SCANLAN, J. O.: 'Properties of the tunnel diode', *Electronic Technol.* 1962, **39**, p. 269

8 Appendix

The matrixes for the blocks of Fig. 8*a* are as follows:

$$[A_1] = \begin{bmatrix} \cos \beta l_1 & jZ_{01} \sin \beta l_1 \\ \dfrac{j \sin \beta l_1}{Z_{01}} & \cos \beta l_1 \end{bmatrix}$$

$$[A_2] = \begin{bmatrix} 1 & 0 \\ Y(\omega) & 1 \end{bmatrix}$$

where $Y(\omega)$ is given by eqn. 2

$$[A_3] = \begin{bmatrix} \cos \beta l_2 & jZ_{02} \sin \beta l_2 \\ \dfrac{j \sin \beta l_2}{Z_{02}} & \cos \beta l_2 \end{bmatrix}$$

$$[A_4] = \begin{bmatrix} 1 & 0 \\ \dfrac{1}{R_s} & 1 \end{bmatrix}$$

$$[A_5] = \begin{bmatrix} \cos \beta l_3 & jZ_{03} \sin \beta l_3 \\ \dfrac{j \sin \beta l_3}{Z_{03}} & \cos \beta l_3 \end{bmatrix}$$

APPENDIX 8.

Short-hop radio-relay systems using tunnel-diode repeaters.

Proceedings I.E.E. 114, 4, pp 435-442, (April 1967).

Short-hop radio-relay systems using tunnel-diode repeaters

D. L. Hedderly, Ph.D., C.Eng., M.I.E.E., J. Hooper, C.Eng., M.I.E.E., and M. K. McPhun, B.Sc., C.Eng., M.I.E.E.

Synopsis

The first part of the paper considers the properties and design requirements of short-hop radio-relay systems; in particular, systems using nondemodulating, nonfrequency-changing repeaters. It is concluded that a system of close-spaced low-power repeaters, mounted on the towers of the high-voltage transmission network, is technically and economically feasible and would be efficient in utilisation of the frequency spectrum. The second part of the paper describes the design of a 2-way u.h.f. tunnel-diode repeater and its performance in the laboratory and in a short experimental link. The u.h.f. repeater was developed to test the practicability of tunnel-diode repeaters for the short-hop system already considered, but, because of the relatively low aerial gain available at u.h.f., its application is restricted to overcoming obstacle loss between two u.h.f. radio terminals. It is finally concluded that tunnel-diode repeaters are feasible for short-hop systems and could find an immediate application in reducing propagation loss over short but difficult radio paths.

List of symbols

α = beam width of aerial, deg
A_c = characteristic hop attenuation, dB
A_e = end-hop attenuation, dB
A_{mn} = attenuation between stations m and n, dB
A_o = overreach attenuation, dB
A_S = system attenuation, dB
C = cost, arbitrary units
C_T = total cost, arbitrary units
D = aerial aperture, m
D_T = aperture of terminal aerial, m
N = noise factor
f_c = carrier frequency, c/s
f_i = interfering frequency, c/s
G = amplifier gain, dB
h = aerial height above ground, m
λ = wavelength, m
S = hop length, m
T = repeater-path delay, s
P = system power level, dBm
I = repeater power input level, dBm
I_P = repeater power input from Pth station, dBm
O = repeater power output level, dBm
O_P = repeater power output toward Pth station, dBm
P_R = terminal-receiver input level, dBm
P_T = terminal-transmitter output level, dBm

1 Introduction

The first use of radio in a point-to-point role was over sea paths which required single-hop paths of extreme length. With the introduction of microwave radio for inland point-to-point communication, it was natural to retain the concept of long hops between terminals, and conventional microwave systems were developed on the basic idea of a narrow radio beam propagated over a clear line-of-sight path of the maximum practicable distance. The major factors limiting the hop distance are the curvature of the Earth and the obstacles upon it; microwave terminals therefore use high ground and/or high aerial towers to obtain a clear line of sight for the required distance. Repeater stations, with similar requirements for aerial height, are necessary when the total distance is much greater than 30 mile.

The foregoing requirements are responsible for the majority of the costs of a microwave system, since they lead, in general, to station sites located at inaccessible points. Therefore, in addition to the cost of the aerial towers, there will be those of access roads, main and standby power supplies and buildings. At terminals situated on low ground, the choice may be between building a very tall aerial tower and installing a coaxial-cable link between the terminal and a favourable radio site. The long-hop principle also leads to less obvious difficulties which, in one way or the other, eventually result in increased costs compared to short-hop designs. These difficulties are:

(*a*) amenity and wayleave problems arising from siting tall aerial towers
(*b*) shortage of suitable sites for planning specific routes
(*c*) fading, which requires greatly increased effective radiated power for an adequate safety factor
(*d*) exclusion of the use of higher frequencies, which are liable to excessive atmospheric attenuation over long hops
(*e*) increased possibilities of interference owing to large effective radiated power and high aerials

On the other hand, a short-hop design would require a large increase in the number of repeaters for a given total distance, and this design principle would show a net advantage only if very cheap and reliable repeaters were available.

In the electricity-supply industry, a part at least of the cost of such repeaters can be discounted, because the close-spaced towers of the high-voltage network offer ready-made sites for the repeater equipment. The high-voltage towers could support a pair of small microwave aerials and amplifying equipment at a modest height of a few metres above the ground. A tunnel-diode amplifier requiring but a few milliwatts of power could furnish the basis of cheap and reliable amplifying equipment. The realisation of a multirepeater short-hop system depends critically upon the installed cost and reliability of a repeater of adequate performance. It was therefore decided to develop and test a repeater of a general type that could be used in such a system.

At the time the work was started, the electricity supply industry was beginning to use single-channel duplex f.m. links in the 470 Mc/s band for the control of v.h.f. base stations of the mobile-radio network and for telemetry purposes. It was foreseen that a cheap nonfrequency-shifting repeater with a self contained power supply, for use in this band, could substantially reduce the cost of links required to span difficult country and would also provide valuable experience in the operation of equipment for use in a future short-hop system; however, it was recognised at the outset that the cost of high-gain aerial systems at 470 Mc/s would prevent the practical utilisation of a repeater suitable for multihop operation at that frequency. The first part of this paper is a feasibility study of short-hop systems which considers some of their more important properties and design requirements. This is followed by a description of the experimental 470 Mc/s repeater, its design and performance. Detailed description of the tunnel-diode amplifiers is given in a companion paper.[1]

Paper 5218 E, first received 18th August and in revised form 1st December 1966
Dr. Hedderly and Mr. Hooper are, and Mr. McPhun was formerly, with the Central Electricity Research Laboratories, Leatherhead, Surrey, England. Mr. McPhun is now with the School of Engineering Science, University of Warwick, Coventry, War., England

2 General properties of repeatered systems in relation to hop length

In this Section, some new basic terms will be defined, and some estimates will be made of the performance of idealised short-hop repeatered systems; these estimates will show the properties in a form which will allow comparison with conventional systems. In particular, the properties of nonfrequency-changing repeaters will be considered. Frequency-division multiplex systems for multichannel speech telephony in which the carrier is frequency-modulated will be used for reference purposes. This is not intended to exclude the possibility of p.c.m. systems, which might well offer advantages when extra bandwidth is available.

2.1 Basic definitions

Attention will be restricted to repeater systems for 2-way transmission in which the same performance is required for both directions. Fig. 1 shows four stations: 1, 2, 3 and 4.

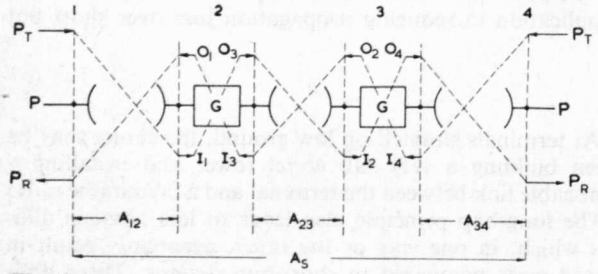


Fig. 1
Basic radio system

The 'hop attenuation' between stations 1 and 2 is given by A_{12}, that between stations 2 and 3 by A_{23} and so on. The hop attenuations include the aerial gains and, by reciprocity, are the same for both directions of transmission. Stations 2 and 3 are repeater stations, and between their aerial terminals are connected bidirectional amplifiers of gain G, given by

$$G = O_3 - I_1 = O_1 - I_3 = O_2 - I_4 = O_4 - I_2$$

where O_P is the repeater output power to the Pth station and I_P is the repeater input power from the Pth station. As defined, the amplifier gain G is an effective gain which includes feeder losses.

Stations 1 and 4 are terminals, and attenuations A_{12}, A_{34} are the 'end-hop attenuations', designated A_e. For equal performance in both directions, the terminal transmitters and receivers are identical and the transmitter output and receiver input power define the 'system attenuation' A_s and the 'system power level' P as follows:

$$A_s = P_T - P_R \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$P = \frac{P_T + P_R}{2} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

The system power provides a common link between input and output power levels at all aerial terminals:

$$P = \frac{O_P + I_P}{2} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

A homogeneous system is one in which all amplifier gains are equal and which has equal hop attenuations, called the 'characteristic hop attenuation' A_c. Referring to Fig. 1 again, a signal arriving at any aerial by a path other than wholly through the repeater chain, is called an 'overreach signal'. It differs from the repeater-path signal in amplitude and time delay. The difference in level between the overreach signal and the repeater-path signal is 'overreach attenuation' A_o. The delay of the repeater-path signal relative to the overreach signal is the 'repeater-path delay' T.

2.2 Power levels in short-hop systems

In a radio system, performance is ensured by first choosing a received-power level P_R sufficiently large compared to the thermal noise level, and increasing this value by a safety factor to allow for fading. In a long-hop conventional system, the fading margin is usually about 30dB. Reduction of hop length has three effects:

(*a*) The hop attenuation is decreased.
(*b*) The total thermal noise is increased because of the increased number of stations.
(*c*) Fading is reduced.

Effect (*a*) permits a reduction in P_T and/or the size of the aerials. Effects (*b*) and (*c*) act in opposition to produce an optimum hop length at which the required P_R is a minimum.

In a homogeneous system of n stations, each with a noise factor N, the overall noise factor is

$$N_n = N + (n-1)(N-1) \quad \ldots\ldots\ldots\ldots \quad (4)$$

and that for a 2-station system is

$$N_2 = N + (N-1) \quad \ldots\ldots\ldots\ldots \quad (5)$$

The increase in noise power when a single-hop link is replaced by an $(n-1)$ hop link is therefore $10 \log (N_n/N_2)$ decibels. This requires an increase in P_R, which is plotted in Fig. 2*a*,

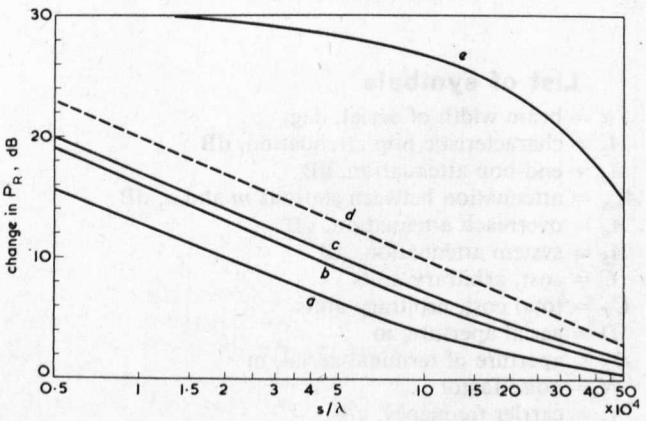


Fig. 2
Effect of hop length S on required receiver input power

a Increase in P_R required relative to one $10^6\lambda$ link, $N = 1$ dB
b Increase in P_R required relative to one $10^6\lambda$ link, $N = 5$ dB
c Increase in P_R required relative to one $10^6\lambda$ link, $N = 15$ dB
d Increase in P_R required relative to large number of $10^6\lambda$ links in tandem
e Reduction in P_R permitted because of the smaller fading range

b and *c* as a function of spacing for three values of N, showing that the increase is not critically dependent upon N. Shown dotted in Fig. 2*d* is the asymptotic value of the increase in P_R if the reference link is a large number of 50km links in tandem. There is some evidence[2] that fading is proportional to the hop length S. Fig. 2*e* also shows the reduction in P_R allowable on this basis when the hop length is reduced. The hop lengths are shown in wavelengths to make some allowance for increased fading at higher frequencies, e.g. more sensitive wave-interference patterns and the onset of rain attenuation. The curve shown is based on a fading margin of 30dB for a 50km path at a wavelength λ of 5cm; it would also represent a 30dB fading margin for a 10km path at $\lambda = 1$cm. The difference between curves *b* and *e* in Fig. 2 gives the net reduction in P_R for an unchanged performance when $N = 5$dB. This value has a broad maximum of about 20dB between 5×10^4 and $25 \times 10^4\lambda$, and the optimum hop length lies between these two distances.

Table 1 shows a comparison between the power levels of a conventional 50km link carrying multichannel telephony according to CCIR recommendations (1959), and two homogeneous close hop systems of 2·5 and 12·5km spacing at $\lambda = 5$cm. All systems are assumed to have a 5dB noise factor, 41dB gain aerials, and a minimum weighted signal/noise ratio of 40dB in the top 3·1kc/s channel during a fade of the appropriate level in one hop. The system power levels are reduced respectively by 25 and 32dB for the 12·5 and 2·5km systems, and the required values of P_T are low enough to be provided by tunnel-diode amplifiers. In a more general short-hop system, the terminal transmitters and receivers will differ from those of the repeaters; therefore the end-hop attenuation A_e will not be equal to A_c, and the terminal power levels P_T and P_R will not be equal to the repeater levels O and I.

Table 1

POWER-LEVEL COMPARISON OF 50, 12·5 AND 2·5 KM SYSTEMS FOR $\lambda = 5$ CM

Number of channels	50 km				12·5 km				2·5 km			
	P_R	A_s	P_T	P	P_R	A_c	P_T	P	P_R	A_c	P_T	P
	dBm	dB	dBm	dBm	dBm	dB	dBm	dBm	dBm	dB	dBm	dBm
240	−55	60	+ 5	−25	−74	48	−26	−50	−74	34	−40	−57
600	−47	60	+13	−17	−66	48	−18	−42	−66	34	−32	−49
960	−43	60	+17	−13	−62	48	−14	−38	−62	34	−28	−45

2.3 Nonlinear distortion

The signals carried by radio-relay systems using frequency modulation suffer distortion if the system group-delay characteristics are nonuniform. Increasing the number of repeaters will add to the distortion unless the group-delay characteristic of the repeater is correspondingly better. This implies that the filters should not be of narrow bandwidth. The filter requirement tends to bias the design towards the higher frequencies, where a large spacing between 'go' and 'return' frequencies is possible. However, distortion from nonuniform delay characteristics is unlikely to be a serious problem, because the overall system characteristics can be compensated or 'equalised'.

The extra reliability required of an increased number of repeaters will cause the design to tend towards simplicity. Much of the complication inherent in conventional systems is caused by the need to change carrier frequencies between hops to avoid overreach interference. However, an overreach signal is only troublesome when the repeater path delay is large and the signal is approximately in phase quadrature with the repeater-path signal. It is therefore of interest to estimate the distortion in a simple nonfrequency-changing system owing to overreach signals.

Medhurst[3] shows that maximum distortion power in the top channel caused by echoes is proportional to the fourth power of the echo delay and to the square of the echo amplitude. Consider a chain of n stations in a straight line. Neglecting the small positive delay of the reflected signals, the distortion noise power tends to a limit for overreach from distant stations, because it increases as $(n-2)^4$ owing to increasing time delay and is inversely proportional to $(n-1)^4$ for overreach signals propagated over a flat earth. This limit is calculated from Medhurst,[3] and the corresponding signal/distortion ratio is shown in Fig. 3 for three channel capacities as a function of the repeater-path delay T.

The total distortion power generated in the system is the sum of the distortion power from each station. The first station 'sees' $n-2$ overreach signals, the second station

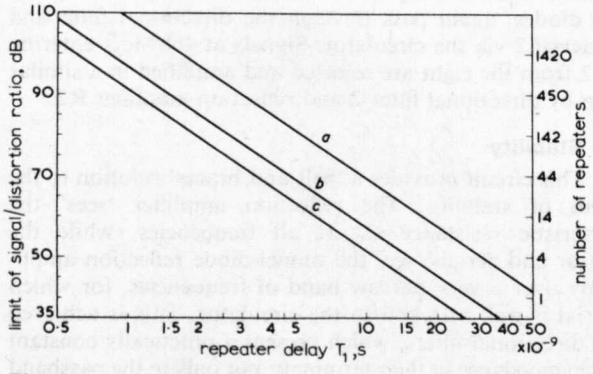


Fig. 3

Limit of signal/distortion ratio for one overreach signal and maximum number of repeaters for 40 dB ratio in top 4 kc/s channel

a 240 channels
b 600 channels
c 960 channels

'sees' $n-3$, and so on. The total distortion power arriving at the first station will therefore not exceed the sum of the arithmetical progression $1+2+3 \ldots n-2$ multiplied by the limiting distortion power. On this basis, Fig. 3 also shows the maximum number of repeaters in a system if the signal/distortion ratio in the top channel is not to be worse than 40 dB. Fig. 3 indicates that, if such short-hop systems are not to be confined to relatively narrow modulation bandwidths and/or a small number of stations, the overall group delay of each repeater must be appreciably less than 20 ns. This will again call for wide filter bandwidths and hence higher frequencies. The designer will also prefer air-spaced coaxial feeders to waveguides and dielectric-spaced cables, horn aerials to dishes, and single- to multistage amplifiers. If these requirements prove too difficult, the distortion may be reduced by dividing the system up into sections with orthogonal polarisation and/or differing frequencies. Zig-zagging of the direction of transmission could also be very effective.

Another source of echo distortion is the feedback loop comprising amplifier input/output and the back-to-back coupling between transmit and receive aerials. The delay around this loop will be no greater than $2T$. In practice, the back-to-back coupling should be of the order of −60 dB for microwave aerials, and, for G not exceeding 30 dB, the distortion would be small compared to that caused by overreach. This source of distortion therefore limits G to a relatively low value for a nonfrequency-changing system.

2.4 Aerials, wavelength, characteristic attenuation and minimum costs

In this Section, it is shown how these design features are interrelated in a homogeneous system and how a suitable choice may be made on the basis of optimum hop length. The generalised relationship between D/λ and S/D for various A_c is shown in Fig. 4 for free-space propagation. The curves

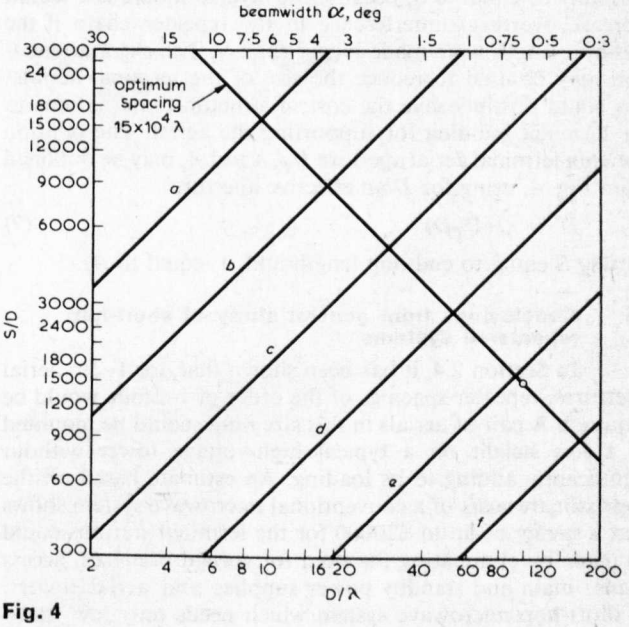


Fig. 4

General relations between major system parameters for free-space propagation

a $A_c = 70$ dB
b $A_c = 60$ dB
c $A_c = 50$ dB
d $A_c = 40$ dB
e $A_c = 30$ dB
f $A_c = 20$ dB

assume aerials of square aperture and 60% efficiency; their beamwidth is shown at the top of the Figure. At right angles to the lines of constant A_c is a line representing the optimum hop length of $15 \times 10^4\lambda$, according to the considerations of Section 2.2. In Section 2.3, it was shown that aerial back-to-back coupling would limit the amplifier gain to relatively low

levels for a nonfrequency-changing system. A_c will therefore also be low, and the design will tend to follow the optimum hop-length line towards the bottom right-hand corner of Fig. 4, i.e. in the direction of large aerial apertures and/or short wavelengths.

Suppose 28dB was tolerable for A_c. The encircled point on Fig. 4 corresponds to this condition and to optimum hop length, and the system design would require $S/D = 1500$ and $D/\lambda = 100$. The repeater hop length and the wavelength are now determined by the choice of aerial size; i.e. for $D = 1$m, $S = 1500$m and $\lambda = 1$cm; for $D = 2$m, $S = 3000$m and $\lambda = 2$cm, etc. The aerial size chosen will depend upon the costs of the aerials and the other costs per repeater. If it is assumed that the cost of the two aerials of a repeater plus a proportion of their installation costs is equal to C_1D^2, and if the other repeater costs are equal to C_2, then, since the number of repeaters required for a given total distance is inversely proportional to D, the total costs of the system will be of the form $C_T = C_1D + C_2/D$. C_T has a minimum when $C_1D^2 = C_2$; i.e. when the aerials and their installation costs equal the other repeater costs.

Referring to Fig. 4 again, a reduction in A_c would require the use of very narrow aerial beam widths, and there would be a limit arising from the costs of holding the aerial steady. However, if the aerials were mounted above the ground so as to profit by an in-phase reflection, A_c could be reduced by up to 6dB using the same beam width. The minimum height for favourable earth reflection is given by

$$h = \sqrt{\left(\frac{S\lambda}{2}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

or, in terms of $S = 15 \times 10^4\lambda$, $h = 200\lambda$. For the two examples given, the aerial heights would therefore be 2 and 4m, respectively. A_c may also be reduced by waveguiding effects between repeaters; e.g. the iterative lens system described by Goubau and Schwering.[4]

In a more general system, A_e may be greater than A_c. For single repeater 2-hop systems, it may often be possible to use this extra attenuation to increase the end-hop length, keeping the terminal aerials the same size. Where the repeater cannot be situated at midpath, the aerials at each terminal will differ in size in order to maintain the same A_e in each hop. In multirepeater systems, the end-hop length should normally be equal to S, because the inverse-square law would increase overreach interference in the repeater chain if the end-hop length were made bigger than S. The extra attenuation may be used to reduce the size of the terminal aerials; this could possibly save the cost of structural alternations to the terminal building for supporting the aerial. The relation between terminal aerial aperture D_T, λ and A_e may be obtained from Fig. 4, using for D an effective aperture

$$D' = \sqrt{(D_TD)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

putting S equal to end-hop length and A_e equal to A_c.

2.5 Conclusions from general study of short-hop repeatered systems

In Section 2.4, it has been shown that, for 1–2m aerial apertures, repeater spacings of the order of 1–2mile would be required. A pair of aerials in this size range could be mounted at a low height on a typical high-voltage tower without significantly adding to its loading. An estimate based on the approximate costs of a conventional microwave system shows that a saving of up to £20000 for the terminal stations could be made by eliminating the need for special premises, access roads, main and standby power supplies and aerial towers. A short-hop microwave system which needs only low transmitter power and low aerials at the terminals could make this saving. The cost of conventional microwave repeaters is up to £600 per mile, and it is evident that the whole of this terminal saving could be recovered if the cost per mile of the short-hop repeaters were no more than, say, £500. Taking the example in the previous Section, this gives a target cost of £500–£1000 per repeater, which comes within the bounds of feasibility.

There are social as well as economic considerations. Short-hop systems radiate very little power, their aerials are close to the ground, and they operate more effectively at higher frequencies not used for conventional radio-relay systems. All these factors tend to make for reduced interference and better utilisation of the radio frequencies, and it is concluded that short-hop systems are feasible on technical and economic grounds and are efficient in their utilisation of the frequency spectrum.

3 470 Mc/s experimental repeater

A low-power nonfrequency-changing repeater may be used as a single repeater between radio terminals, in order to overcome the transmission loss due to an obstacle. This is a different role from the one played in multihop systems, but similar savings in terminal costs and a reduction in transmitter power can be achieved in many cases. In this role, the repeater may be conveniently thought of as a passive repeater consisting of a pair of back-to-back aerials. For example, aerials of 20dB gain for 470 Mc/s working are readily obtainable; insertion of a 20dB straight-through amplifier between two such aerials effectively converts them to a pair of 30dB gain aerials connected back to back. This will not only increase the permissible distance between terminals and repeater, but it will also increase the overreach attenuation for a particular obstacle. Both of these factors increase considerably the number of situations in which such a repeater is feasible.

The development work was aimed at a low-cost amplifier of 20dB gain for the 450–470 Mc/s band. Only tunnel diodes were considered as active elements, because, at the time the work was started, they alone had sufficiently high cut-off frequencies and low power consumptions to meet the requirements for the more long-term objective of multihop systems. Two designs of amplifiers were produced which used the same tunnel-diode-amplifier module in two different configurations of filters and circulators.

3.1 Type I amplifier

Amplifier stability was the major initial problem to be faced in the design of the repeater. Narrowband aerials were to be used on the output and input of the amplifiers (narrowband filters were required, the only circulators available at 470 Mc/s had a narrow bandwidth), and it was proposed to use common transmit and receive aerials for the two directions of transmission. All these factors increased the likelihood of spurious oscillations from the tunnel diodes, either at the signal or some other frequency. After describing the principle of the repeater, we shall see how stability was achieved.

The diagram of the type I amplifier is shown in Fig. 5, in which the reflection amplifiers are those described in a companion paper.[1] The action is as follows: signals at 454 Mc/s enter from the aerial at the left, pass round the circulator as shown, and reach reflection amplifier R1 via directional filter 1. They are reflected and amplified by the tunnel diodes, again pass through the directional filter and reach aerial 2 via the circulator. Signals at 468 Mc/s entering aerial 2 from the right are selected and amplified in a similar fashion by directional filter 2 and reflection amplifier R2.

3.1.1 Stability

This circuit provides a 'belt and braces' solution to the problem of stability. The reflection amplifier 'sees' the characteristic resistance Z_0 at all frequencies, while the circulator and aerials 'see' the tunnel-diode reflection amplifier only over a very narrow band of frequencies, for which the aerial is well matched to the circulator. This is achieved by the directional filters, which present a practically constant resistive impedance at their terminals, not only in the passband but also in the stopband. Thus any signal at 454 Mc/s which is reflected from aerial 2 will be passed via the circulator into directional filter 2, where it is absorbed. The bandwidth of the directional filters was made much narrower than that of the aerials and the circulator.

3.1.2 Directional filters

Fig. 6 shows the action of the directional filters. A 3dB directional coupler or 'hybrid' has its two coupled ports terminated in identical bandstop filters. Signals entering at A are split equally and impinge on the bandstop filters. In the stopband of these filters the signals are totally reflected,

recombine in the hybrid, and emerge at B. This is the passband of the directional filter. Outside this band the signals pass through the bandstop filters and are dissipated in their terminations. No signal is reflected back to A at any frequency. Naturally, in practice it is not possible to maintain

3.1.3 Laboratory performance

The gain/frequency response of the type I amplifier was tested at room temperature using equipment connected as described in the companion paper.[1] It was found that two main types of gain/frequency response of the repeater could

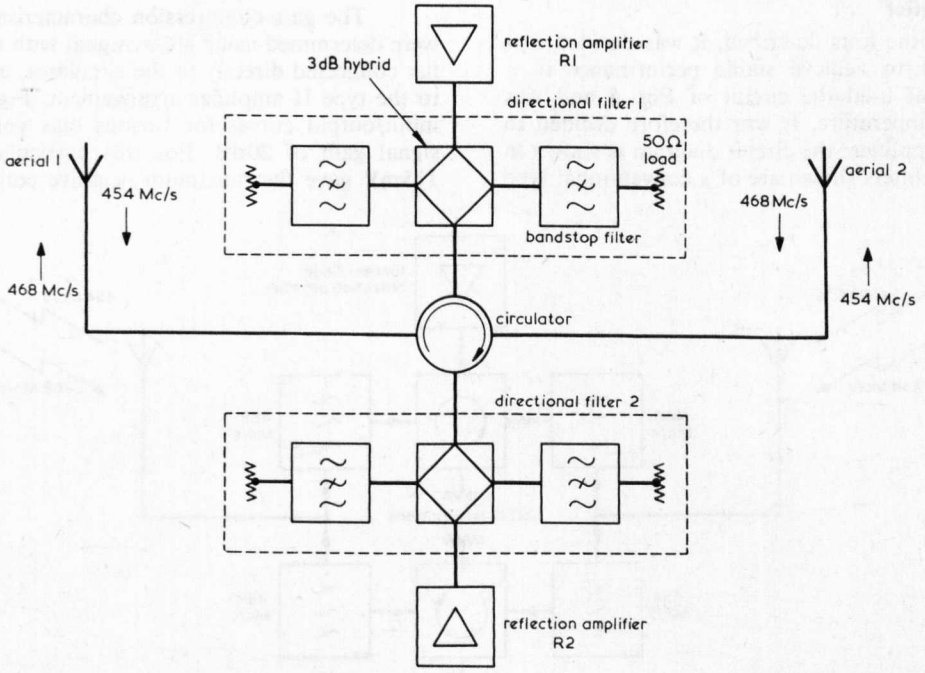


Fig. 5
Diagram of type I tunnel-diode amplifier

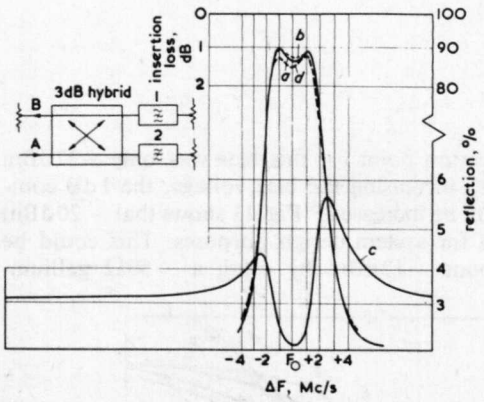


Fig. 6
Schematic diagram and performance of directional filters

a Reflection of filter 1
b Reflection of filter 2
c Reflection at A
d Insertion loss A to B

be obtained, depending on the adjustment of the constant-impedance bandpass filters. The first type of response consisted of a single-gain peak in each direction of transmission; an example of this type of response is shown in Fig. 7. These results were obtained with a 3dB filter bandwidth of about 3 Mc/s and a minimum filter loss of 2dB.

The second type of response is shown in Fig. 8, and consists of a double-peaked response with a 3dB bandwidth of about 3 Mc/s. In order to achieve this response, the filter had to be adjusted to give a bandwidth of about 6 Mc/s, and consequently had a greater midband loss of approximately 6dB for a signal passing through the filters and back again. In the course of making these measurements, it was found that the directional filters were rather critical in adjustment owing to their narrow bandwidth and also because they were arranged in a bridge structure.

Tests were carried out to determine the effect of temperature variations on the gain/frequency characteristic with the filters adjusted to give both types of response described. In both cases the gain of the 454 Mc/s amplifier increased when the

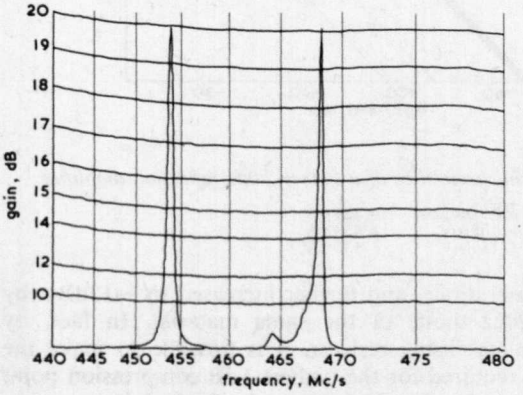


Fig. 7
Gain of type I amplifiers showing single-peaked response

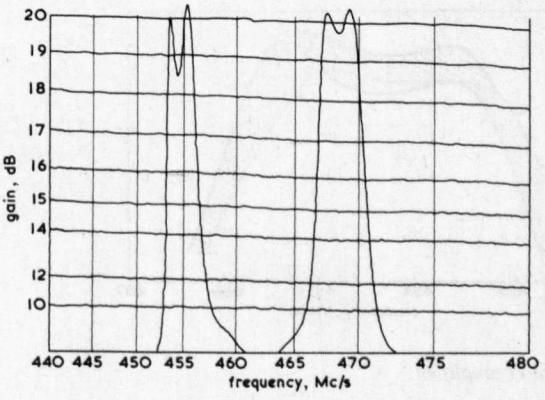


Fig. 8
Gain of type I amplifiers showing double-peaked response

a perfect balance in the filter, and the curves of Fig. 6 show what was achieved. The bandstop filters consisted of two coaxial cavities, spaced $\lambda/4$ apart, each with a loose capacitive coupling to a coaxial transmission line.

temperature was lowered from 20°C to −10°C; in the case of the double-peaked response, the amplifier broke into oscillation. By changing the tunnel-diode bias voltage, and thus the tunnel-diode negative resistance, the gain at any

particular frequency could be restored to its original value. However, the shape of the gain/frequency characteristic was distorted, severely so in the case of the double-peaked response. The gain/frequency characteristic of the 468 Mc/s amplifier was also distorted by a similar temperature change.

3.2 Type II amplifier

As a result of the tests described, it was decided that it would be difficult to achieve stable performance in a practical repeater that used the circuit of Fig. 5 and was subject to varying temperature. It was therefore decided to construct a type II amplifier; the circuit diagram is shown in Fig. 9. The diplexing filters shown are of a conventional type

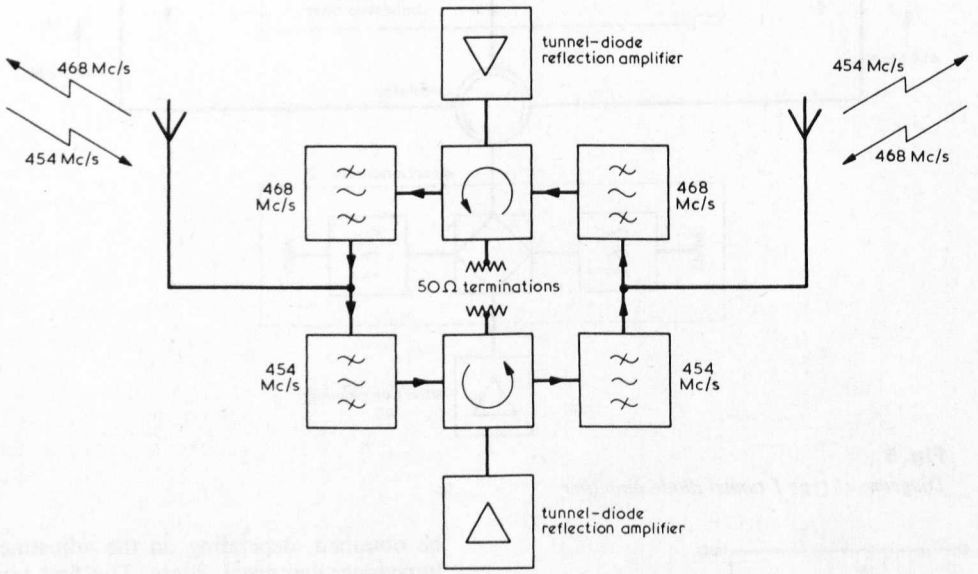


Fig. 9
Type II amplifier configuration

and are separated from the two reflection amplifiers by circulators. The reflection amplifiers are thus isolated from changes in the filter v.s.w.r. by the reverse loss of the circulators, i.e. at least 20 dB, and the amplifier gain is now determined by the reflection-amplifier characteristic and the circulator only. The difficulty with this arrangement is that the reflection amplifiers must be designed to be stable when connected directly to the circulators; however, this has proved to be possible.[1]

3.2.1 Gain/frequency response

Some typical gain/frequency response curves of the type II amplifier for various temperatures are shown in Fig. 10. These curves were measured using a constant bias

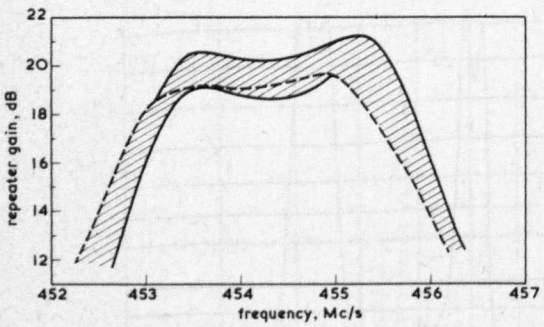


Fig. 10
Gain of type II amplifier

--- gain at 20° C
Shaded area shows total gain variation between ±20° C

voltage, and it can be seen that the total gain variation at the centre frequency is less than 2 dB for temperatures between −20° C and +20° C. This was achieved without special selection of the circulators or temperature compensation of the filters, and suggests that a total gain variation of 2 dB for a temperature range of −20° C to +50° C could be achieved in future models of the equipment. Since in practice the bias voltage would also vary with temperature, it should be possible to use this variation to help keep the gain constant.

3.2.2 Maximum power output and gain-compression point

The gain-compression characteristics of the amplifier were determined using a c.w. signal with the reflection amplifier connected directly to the circulator, as this corresponded to the type II amplifier arrangement. Fig. 11 shows a set of input/output curves for various bias voltages and a small-signal gain of 20 dB. For this particular diode, a bias of 115 mV gave the maximum negative conductance; however, the 1 dB compression point for this bias was only −27 dBm at the output. By increasing the bias voltage, the 1 dB compression point can be increased;[5] Fig. 11 shows that −20 dBm can be assumed for system design purposes. This could be increased to about −17 dBm by using a −50Ω gallium-

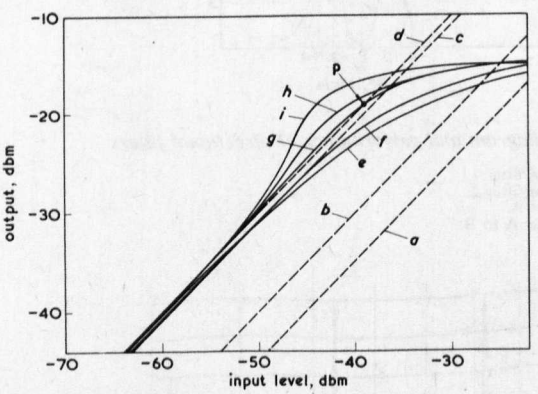


Fig. 11
Input/output characteristics of a tunnel-diode reflection amplifier

a 5 dB gain	*d* 20 dB gain	*g* 125 mV
b 10 dB gain	*e* 105 mV	*h* 130 mV
c 19 dB gain	*f* 115 mV	*i* 135 mV

arsenide tunnel diode, and further increased to −13 dBm by using a −20Ω diode of the same material. In fact, by selecting the stabilising resistor, it is possible to make the bias voltage required for the highest 1 dB compression point coincident with that for minimum noise factor.[6]

3.2.3 Distortion and crosstalk with frequency-modulated carriers

The principal use of the u.h.f. repeater will be in links using narrow-deviation f.m. carriers, and tests were undertaken to determine the performance of the repeater when amplifying such signals. The first test employed an r.f.

carrier frequency-modulated by audio frequencies of 3kc/s and 10kc/s each separately set to give 10kc/s deviation; this was passed through the repeater while varying the r.f. input level. The output of the repeater was passed through a variable attenuator, adjusted to give a constant input level to a conventional f.m. receiver. The receiver output was measured by an audio-frequency wave analyser.

Without the repeater in circuit, the intermodulation product at 23kc/s was 55dB down on the level of either the 10kc/s or the 3kc/s tone. With the repeater in circuit there was no observable increase of this intermodulation product until the input to the tunnel-diode amplifier module was −9·4dBm; i.e. a level above any likely to be encountered in practice. In order to check this result, an attempt was made to measure the amplitude-to-phase conversion of the tunnel-diode amplifier. This proved difficult owing to the low level of the signals involved and the lack of an attenuator of known phase characteristics. An estimate was obtained by passing a signal through the amplifier and comparing its phase with a reference signal. This test showed no discernible change of phase shift through the amplifier until its input signal reached −6dBm; a change of the order of 1° per dB was then observed.

Since the bandwidth of the repeater is considerably wider than the channel spacing in the 470Mc/s band, there is the possibility that an unwanted signal will enter the repeater, while being outside the bandwidth of the receiver at the far end of the link, and cause interference with a wanted signal. In order to check this, an interfering f.m. carrier f_i modulated with a 1kc/s tone at 10kc/s deviation was set 1Mc/s away from the receiver frequency, and was passed through the amplifier together with an unmodulated carrier f_c at the receiver centre frequency. The wanted carrier f_c was set to a level where it just began to compress the gain of the tunnel-diode repeater, and the level of the interfering carrier f_i was increased while looking for any sign of a 1kc/s tone in the receiver output. Although the interfering carrier was increased to a level where it had compressed the gain of the amplifier to 0dB, no 1kc/s tone was observed. From this it can be concluded that, while an interfering f.m. carrier entering the repeater may reduce its gain, it will not cause any intelligible crosstalk.

3.2.4 Distortion and crosstalk with amplitude-modulated carriers

The gain-compression curves of Fig. 11 suggest that the tunnel-diode amplifier will distort a.m. carriers if these are of sufficient magnitude to affect its gain and also that intelligible crosstalk will occur from a strong signal to a weak one. Tests showed that both these effects did occur. The distortion of a.m. carriers was tested by applying an r.f. carrier, 50% amplitude-modulated with a 1kc/s tone, to the tunnel-diode amplifier and feeding its output via a variable attenuator to an a.m. receiver. The r.f. input to the amplifier was increased, and the attenuator was adjusted to keep the receiver input constant. It was found that the 1kc/s output from the receiver decreased approximately in agreement with the drop in r.f. gain of the repeater to a point where the r.f. gain had fallen from 20dB to about 5dB; from then on, the 1kc/s signal fell rapidly as the repeater gain was compressed to zero.

In order to determine the crosstalk from an interfering a.m. carrier, an r.f. carrier f_c, 50% amplitude-modulated with a 1kc/s tone, was passed through the amplifier and fed to an a.m. receiver; the resultant 1kc/s audio output from the receiver was used as a reference. The r.f. carrier f_c was set to −50dBm, which was just low enough not to affect the tunnel-diode amplifier gain.

The modulation was then removed from the carrier f_c, and a second similarly modulated interfering carrier f_i set 1Mc/s away from f_c was applied to the tunnel-diode amplifier, together with the unmodulated carrier f_c. The interfering carrier f_i was then increased while observing the audio output, and the level of the resultant 1kc/s tone was measured. During these tests, the r.f. input to the a.m. receiver was maintained at a constant level by adjusting a variable attenuator between it and the tunnel-diode amplifier. The results are shown in Fig. 12, where it can be seen that severe crosstalk occurred for quite low interfering signals. It is concluded that the repeater is unsuitable for an a.m. system.

3.3 Experimental link

An experimental link was installed at CERL to demonstrate the operation of the repeater under conditions approximating to those in practice and to measure the near- and far-field coupling of typical aerial systems. The terminal sites were separated by about 700m, and the line of sight was obscured by a rise in the ground topped by a laboratory building, making an obstruction approximately 30m above the terminals. The repeater was mounted on steel scaffolding so as to overlook the building and to have a clear view of both terminals. The terminal equipment was a commercial 470Mc/s, single-channel, duplex-equipment type. The maximum transmitter output power was 2W and the frequencies were 454·225Mc/s and 468·225Mc/s, respectively, from terminals 1 and 3. During the tests the repeater, which requires a continuous d.c. power supply of 15mW, was installed in a pressurised box and powered from mercury cells. The short length of the link was very convenient for experimental purposes; power levels at the repeater were adjusted to correspond to longer distances by inserting attenuation between the terminal aerials and terminal equipment. The repeater increased the strength of the received signal by more than 24dB for both directions of transmission.

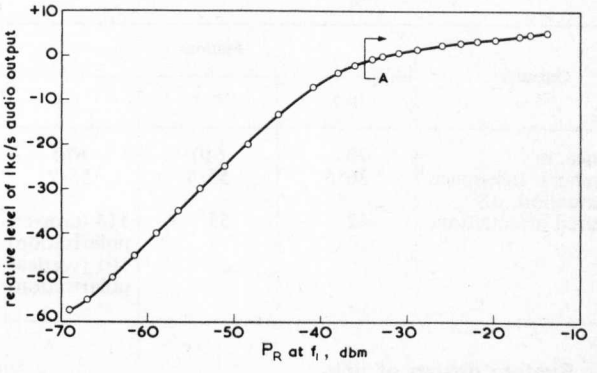


Fig. 12

Relation between level of interfering signal and resultant audio output

A.F. distortion evident in region A

3.3.1 Aerials and radio paths

The repeater aerials were stacked and bayed 14-element Yagi aerials with curtain reflectors, each of 20dB measured gain relative to an isotropic aerial, and those of the terminal were single Yagi aerials of 6dB measured gain. The repeater aerials were installed back to back, separated by about 2·5m between the curtain reflectors, and the attenuation between the terminals of the two aerials was measured. The measurements were made over 440–480Mc/s for a range of horizontal angles between the aerials of 180° to 140° and for different polarisations. The minimum attenuation recorded is shown in Table 2.

Table 2

MINIMUM ATTENUATION BETWEEN TERMINALS OF BACK-TO-BACK AERIALS

Condition	Attenuation
	dB
Both vertically polarised .	50
Both horizontally polarised.	42
Cross-polarised . . .	>55 (limit of measurement)

The 42dB back-to-back coupling provided a very adequate margin against instability using a 20dB amplifier. In the final installation the polarisation between stations 1 and 2 was vertical, and that between 2 and 3 was horizontal. This was not done to reduce back-to-back aerial coupling, but to obtain the maximum overreach attenuation for demonstra-

tion purposes. Table 3 shows the hop lengths together with measured and theoretical path attenuations. The measured hop attenuations were made at the frequency of the associated terminal transmitters and the theoretical calculations at 460 Mc/s.

Table 3

HOP LENGTHS AND ATTENUATION

Quantity	Stations		
	1–2	2–3	1–3
Distance, m . . .	70	610	670
Theoretical free-space attenuation, dB	36·6	55·4	56·2
Measured attenuation, dB	42	55	114 (crossed polarisation) 101 (vertical polarisation)

3.3.2 System design of link

A Post Office specification specifies a minimum output signal/noise ratio of 40 dB for 450–470 Mc/s angle-modulated point-to-point systems. The receiver input power specified for this signal/noise ratio is −95 dBm, and this is therefore the minimum allowable P_R for the system. The minimum P_R and the maximum O (repeater output overload level) fix the maximum A_e for the system. For a given repeater gain, the maximum A_e in turn fixes the maximum P_T allowable before the repeater overloads. Maximum system power levels and system attenuations can then be derived. The fixed values and bounds to the actual system design are listed below.

Amplifier gain $G = 20$ dB
Overload output level $O = -28$ dBm (allowing 1 dB for losses between aerial and amplifier)
Maximum end-hop attenuation $A_e = 67$ dB
Maximum transmitter power $P_T = +19$ dBm
Maximum system power level $P = -38$ dBm
Maximum system attenuation $A_S = 114$ dB

The end-hop attenuations shown in Table 3 require equalising for a symmetrical system. For the experimental link, A_{12} was increased to 55 dB by adding 13 dB in the terminal-aerial feeder; in a working link, the end-hop attenuations would be equalised by an appropriate choice of terminal aerials.

The link was then equivalent to a 2-hop system, with each hop of length 610 m and a characteristic attenuation of 55 dB. The latter attenuation was 12 dB less than the maximum, and P_R was correspondingly greater than the minimum, thus providing a fading margin of 12 dB. To suit the lower attenuation, P_T was reduced to 12 dB below the maximum P_T, i.e. to $P_T = 7$ dBm. A_S was 90 dB, and, since the hop attenuation A_{13} had also been increased by 13 dB to 127 dB, the overreach attenuation was 37 dB. The second column of Table 4 summarises the system characteristics for operation at the normal overload point.

Using the foregoing power levels, the hop length of 610 m could be extended to 1·5 km by the use of 14 dB* gain terminal aerials, and to 3 km with 20 dB aerials. By sacrificing the fading margin the hop length is then extendable to a limit of 12 km. Further extension by increasing aerial gain could prove expensive, and it is better to increase the overload point of the repeater. The third column of Table 4 shows the system characteristics with the amplifier operated at a higher bias level (Point P in Fig. 11), giving O (max) = −20 dBm and $G = 20$ dB. The additional 8 dB of fading margin would then permit a maximum hop length of 30 km.

* 14 dB gain relative to an isotropic aerial is the minimum allowed for 470 Mc/s links in Post Office specifications

Table 4

SUMMARY OF EXPERIMENTAL 470 M/CS SYSTEM CHARACTERISTICS

Quantity	Value	
	O (max) = −28 dBm	O (max) = −20 dBm
A_e (max)	67 dB	75 dB
$A_{12} = A_{23} = A_e = A_c$	55 dB	55 dB
$P = P$ (max)	−38 dBm	−30 dBm
P_R	−83 dBm	−75 dBm
P_T	7 dBm	15 dBm
$O_3 = O_1 = O$ (max)	−28 dBm	−20 dBm
$I_3 = I_1$	−48 dBm	−40 dBm
A_S	90 dB	90 dB
A_{13}	127 dB	127 dB
A_0	37 dB	37 dB
Fading margin	12 dB	20 dB

4 Conclusions

A 2-way, nonfrequency-changing, u.h.f. radio repeater, using tunnel diodes, has been developed, and its use was demonstrated in a short experimental link. When redesigned for frequencies in the 15–30 Gc/s band, it would be suitable for a future short-hop radio-relay system, in which the repeaters were mounted at a modest height on the towers of the high-voltage transmission network. The tunnel-diode radio repeater has an immediate application in reducing transmission loss for short but difficult radio paths at u.h.f. and higher frequencies.

5 Acknowledgments

The authors wish to express their thanks to R. E. Martin for his support and encouragement, P. H. Hawkins, G. R. Francis and H. E. Spencer for carrying out measurements and to G. Berg for the mechanical construction of the repeaters. They also wish to thank O. Kiertzner of Storno Ltd. for the loan of 470 Mc/s terminal equipment. This work was carried out at the Central Electricity Research Laboratories, and is published by permission of the Central Electricity Generating Board.

6 References

1 McPHUN, M. K.: 'U.H.F. tunnel-diode amplifier.' (see pp. 428–434)
2 JONES, D. G.: 'Operational experience with wideband radio links'. To be published as CCIR Document V/11–E, 1965
3 MEDHURST, R. G.: 'Echo-distortion in frequency modulation', *Radio Electronic Engr.*, 1959, **7**, p. 253
4 GOUBAU, G., and SCHWERING, F.: 'On the guided propagation of electro-magnetic wave beams', *IRE Trans.*, **AP–9**, 1961, **3**, p. 248
5 HAMASAKI, J.: 'A low-noise and wide-band Esaki diode amplifier with a comparatively high negative conductance diode at 1·39 Gc/s', *IEEE Trans.*, 1965, **MTT–13**, p. 213
6 McPHUN, M. K.: 'A tunable X band tunnel diode amplifier', Proceedings of the 6th international conference on microwave and optical generation and amplification, Cambridge, England, Sept. 1966

APPENDIX 9.

A tunable X-band tunnel diode amplifier.

Proceedings of the 6th international conference on Microwave and Optical Generation and Amplification, Cambridge, September 1966.
I.E.E. Conference Publication No. 27, pp 292-296.

A TUNABLE X-BAND TUNNEL DIODE AMPLIFIER

M. K. McPhun *

INTRODUCTION When the resistive cut-off frequency (fr) is less than the self resonant frequency (fx) of a tunnel diode, its terminal admittance is well behaved. The diode may be approximated at frequencies up to fr/3 by a negative conductance shunted by a capacitance, both frequency invariant. However, for the diode with $fx < fr$ the terminal conductance and effective capacitance may both vary rapidly with frequency, and moreover, small changes in the diode parameters may give disproportionately large changes in these values.

In spite of the minutely packaged diodes now available we must accept diodes having $fx < fr$ for use in X-band. Some adjustment to the characteristics of each diode is needed. In this amplifier these adjustments are simple to make.

DESCRIPTION The amplifier is of the circulator-coupled reflection type. A 3 port waveguide circulator having isolation greater than 30 dB over the range 10.7 to 11.7 GHz is used. The reflection amplifier is made in coaxial line and mounted on an integral transformer to waveguide; Fig. 1 shows a sketch of the amplifier and its circuit diagram.

The tunnel diode is mounted directly across the transmission line and is preceded by a quarter-wave transformer with a variable characteristic impedance. This is used as the gain control. The variation is achieved, as shown in Fig. 5, by moving one part of the coaxial outer conductor with respect to the inner conductor. By suitably shaping the movable part we obtain a range of characteristic impedance Z_{01} from 9 to 35 ohms. The circuit of Fig. 1a is completed by a diode tuning inductance, and a stabilizing circuit. The latter consists of a 20 ohm resistance shunted by a narrow-band series resonant circuit consisting of C_3 and a short circuited line Z_{03}. At resonance the 20 ohm resistance is effectively shorted, and the diode capacitance is tuned by the diode tuning inductance. Off resonance the circuit is damped by the 20 ohm resistor. The bandwidth of the stabilizing circuit is much narrower than that of the diode and inductance, so the amplifier operating frequency can be tuned by varying C_3. The amplifier bandwidth is determined by the bandwidth of the stabilizing circuit, which may be changed by adjusting the length L_3.

As shown in Fig. 1b the 20 ohm resistor consists of a length of 20 ohm coaxial line of 0.125 inch O.D. terminated in a tapered poly-iron load. The D.C. bias resistance R_b is provided by a carbon resistor, and can differ from 20 ohms if required. The series resonant circuit is also made in 0.125 inch O.D. coax with Z_{03} equal to 100 ohms. C_3, which is of order 0.05 pF, consists of capacity between the hat and the inner of the 20 ohm coaxial line. Stray capacitance between the hat and the outer is 0.04 pF and is independent of the position of the inner. The photograph of Fig. 2 shows a plan view of the amplifier sectioned through the centre of the coaxial lines.

---

* Mullard Research Laboratories, Redhill, Surrey.

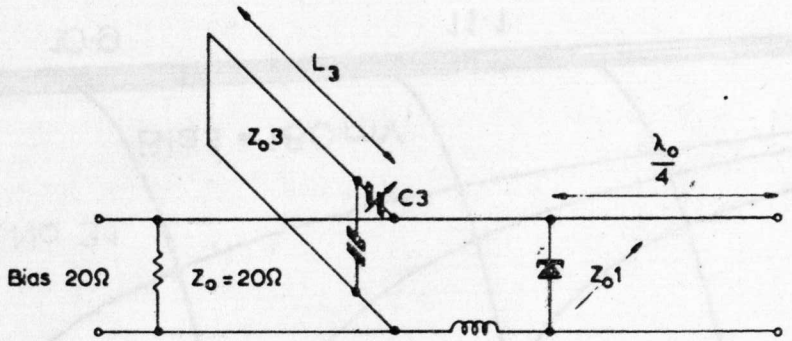
RESULTS A typical set of gain-frequency responses, as shown in Fig. 3, can be obtained by tuning the amplifier using C_3, then setting the gain using Z_{01}. These curves show that the amplifier can be tuned from 10.7 GHz to 11.7 GHz with a $\frac{1}{2}$ dB bandwidth of 30 MHz at 15 dB gain. The air gap between the hat and the inner of the 20 ohm line is shown for each curve, as is also the value of Z_{01}. Note that for a constant gain of 15 dB the value of Z_{01} must be reduced as the resonant frequency is increased. This is as expected; it means that the magnitude of the real part of diode terminal admittance is increasing with frequency.

The set of curves of Fig. 3 can be reproduced for a wide range of values of diode bias; different values of Z_{01} and slightly different C_3 values being the only changes required. Thus the bias can be set to the value required to give the lowest noise factor. We have obtained an overall noise factor of less than 6 dB including contribution from the following receiver. The noise factor of the reflection amplifier is 5.5 dB compared with 4.8 dB predicted using the parameters given by the diode manufacturer.

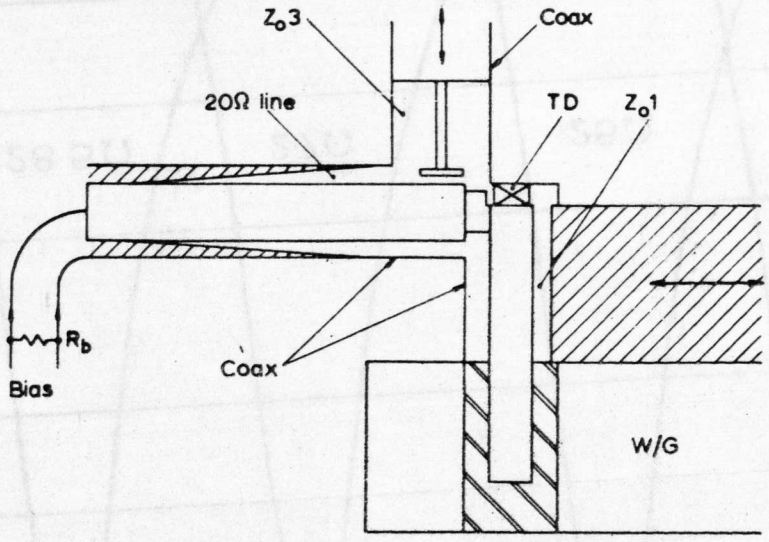
Diode bias can also be selected to give the best saturation characteristic[1], but this bias, in general, will not be that required for minimum noise. However, we have found that the 1 dB compression point (input power in dBm to give 1 dB of gain compression) is also a function of the internal resistance of the bias supply. Fig. 4 illustrates this; it shows saturation characteristics as a function of bias voltage with constant bias resistance, and also as a function of resistance with constant bias voltage. Thus we can select bias voltage to give lowest noise factor, then choose bias resistance to give the required saturation curve. We obtain a 1 dB compression point of -33 dBm with a 2 mA TD406 diode.

CONCLUSIONS The amplifier may be set up in a simple logical way within the 10.7 to 11.7 GHz band to give simultaneously a 30 MHz bandwidth to the 0.5 dB points, a 6 dB overall noise factor and a 1 dB compression point of -33 dBm.

Reference 1 Hamasaki J. "A low noise and wide-band Esaki diode amplifier with a comparatively high negative conductance diode at 1.3 Gc/s". I.E.E.E. Trans. 1965, MTT-13, p.213.



(a) Circuit diagram



(b) Sketch

FIG.1. REFLECTION AMPLIFIER

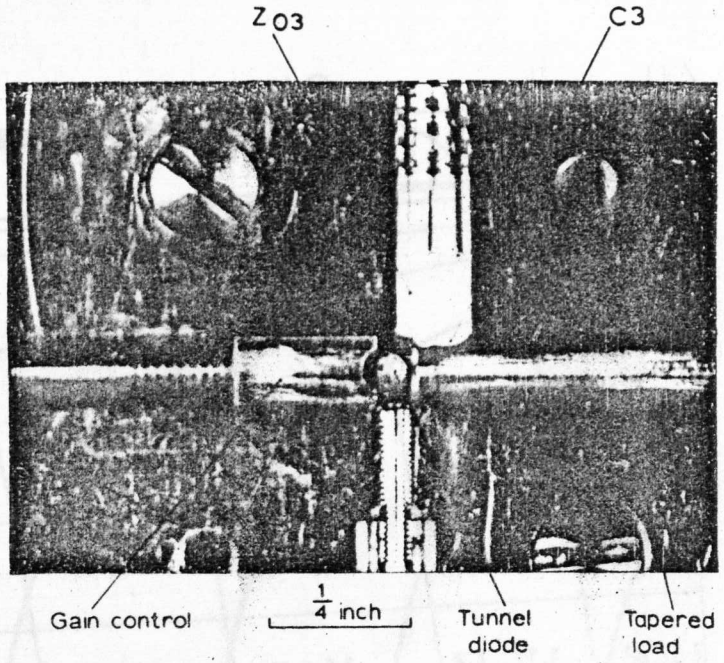


FIG. 2.

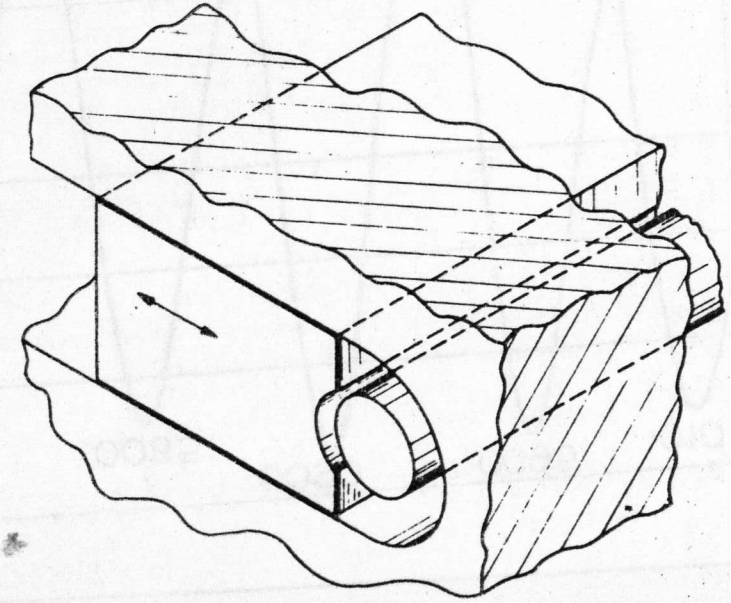


FIG.5. VARIABLE IMPEDANCE TRANSMISSION LINE

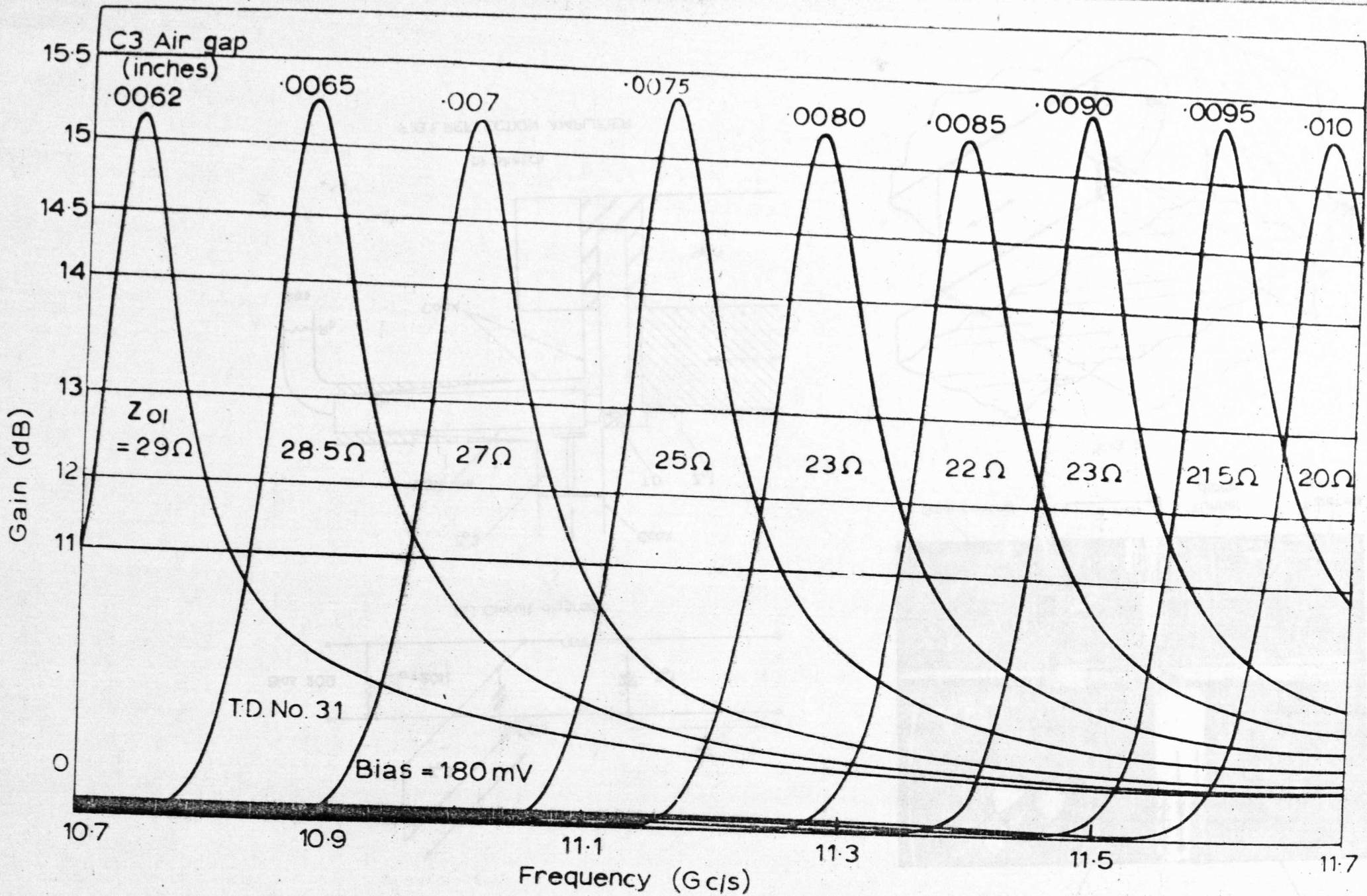
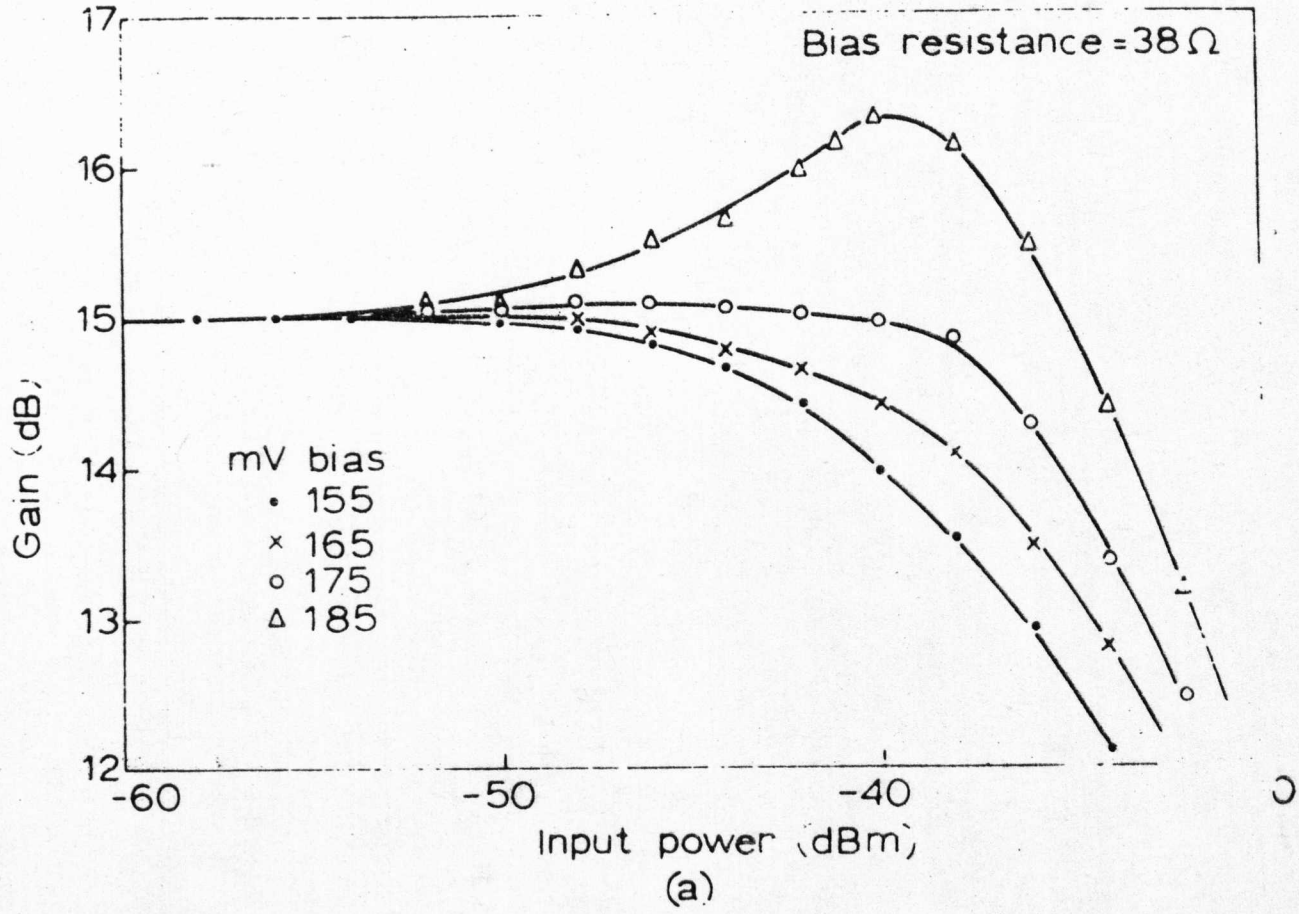


FIG 3. GAIN FREQUENCY RESPONSES AS AMPLIFIER IS TUNED ACROSS THE BAND. GAIN WAS RESET USING Z_{oi} FOR EACH CURVE.



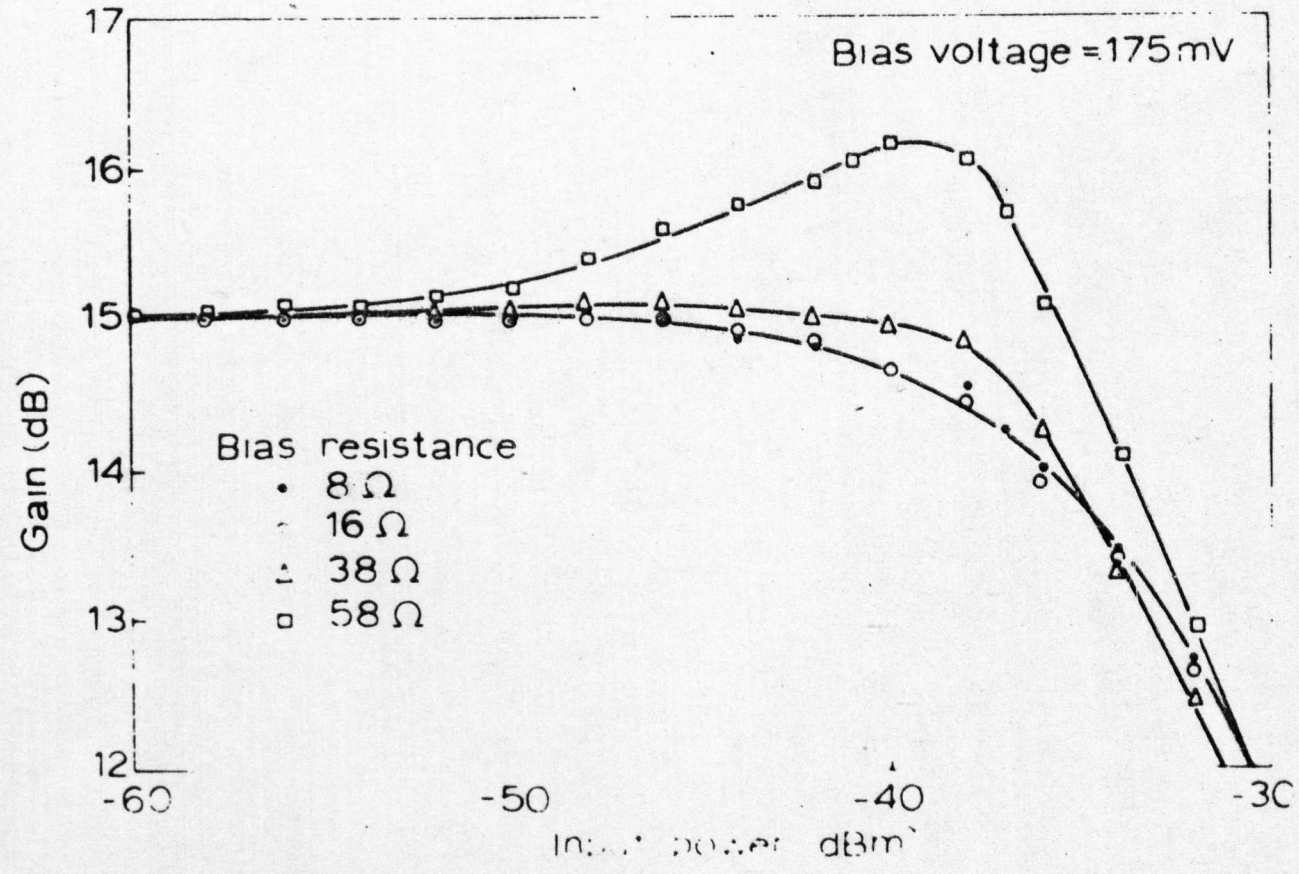


FIG 4 SATURATION CURVES FOR AMPLIFIER
a Set up with bias of 175 mV
b. Bias resistance constant at 38 Ω

APPENDIX 10.

Comparison of TEM with waveguide-below-cutoff resonators.

Electronics Letters 5, 18, pp 425-426 (4th Sept. 1961).

Reprinted from *Electronics Letters*, Vol. 5, No. 18, 4th September 1969

COMPARISON OF TEM WITH WAVEGUIDE-BELOW-CUTOFF RESONATORS

An alternative way of looking at waveguide-below-cutoff resonators for tuning active devices is considered. By treating the circuit as a TEM transmission-line resonator, a simple design technique is possible. Tuning is shown to be accomplished by varying the characteristic impedance of the TEM line. A simple way of obtaining a large increase in the mechanical tuning range is suggested.

The 'evanescent-mode oscillators' described by Ivanek *et al.*[1] employed lengths of waveguide operated below the cutoff frequency. Effects of mechanical tuning were described, utilising movable short circuits in the waveguide and a tuning plunger in the sidewall. The purpose of this letter is to show that the action of these circuits can be simply described in terms of a TEM-mode resonator; the tuning is accomplished by varying the characteristic impedance of the TEM transmission line.

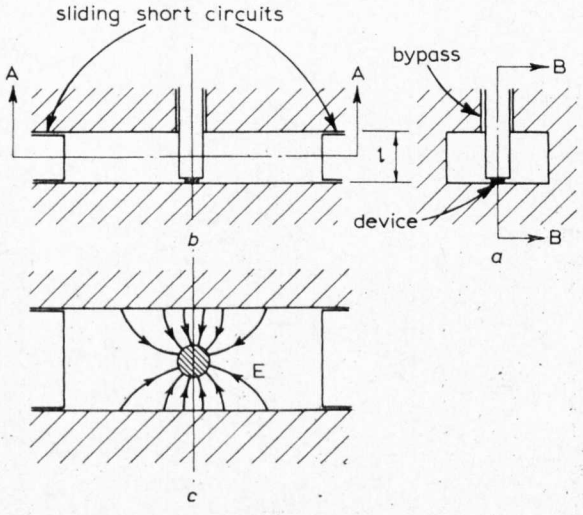


Fig. 1 *Device mounted in waveguide below cutoff*

a Side elevation
b Front elevation sectioned on B–B
c Plan sectioned on A–A

Fig. 1 shows three views of a semiconductor device mounted in a waveguide below cutoff, where Fig. 1*b* is similar to Ivanek's Fig. 1A. The coupling loop has been omitted for simplicity, as we are concerned here only with the conditions for resonance, which need not be affected by the coupling loop. The plan view (Fig. 1*c*) appears as a cross-section through a TEM transmission line. Without the sliding short circuits, it would appear as a slab line, and with them, it has the form of a coaxial line with a rectangular outer. The characteristic impedance Z_0 of this TEM line is determined by the waveguide width, the position of the sliding short circuits and the diameter of the diode support post. These are all important, whereas only the waveguide width determines its cutoff frequency.

The equivalent circuit of the oscillator is simply as Fig. 2*a*, where the variable Z_0 controls the tuning. The admittance seen by the device is $Y = \dfrac{-j}{Z_0} \cot \dfrac{\omega l}{c}$, where $c = 3 \times 10^8$ m/s.

For resonance we have

$$\cot \frac{\omega l}{c} = B_d Z_0 = \omega C_d Z_0 \quad \ldots\ldots\ldots \quad (1)$$

where C_d is the device capacitance. Thus Z_0 determines the resonant frequency. Note also that l is the waveguide height, and this is also an important factor.

The use of TEM lines with variable characteristic impedances as tuning elements is not new.[2] They have been used to control the resonant frequency[3] and the gain[3,4] of tunnel-diode amplifiers. As shown by the lines of electric field sketched in Fig. 1*c*, the field is mainly concentrated in the vicinity of the device support post as surmised by Ivanek *et al.*[1] However, evanescent modes do not enter the picture unless the sliding short circuits are removed. In that case, the waveguide must be below cutoff to prevent propagation sideways out of the slab line. This is, of course, a normal criterion for use of a slab line.

The Z_0 obtained is a very nonlinear function of the position of the sliding short circuits. This is entirely consistent with the tuning curve given by Ivanek *et al.*[1] in their Fig. 3. Their use of a tuning plunger in the sidewall of the waveguide is just another way of varying Z_0. If the plunger does not cover the full height of the waveguide, the equivalent circuit becomes more complicated, as shown in Fig. 2*b*.

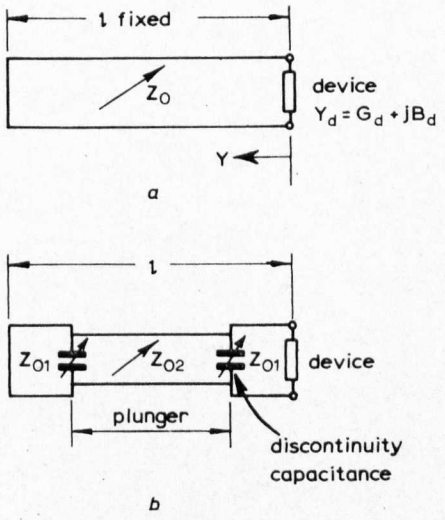


Fig. 2 *Equivalent circuits*

a Tuning by sliding short circuits
b Tuning by plunger in sidewall

The virtue of considering the circuit as a TEM resonator is that it renders the circuit designable. Eqn. 1 and the determination of Z_0 is all that is required. For the slab line, Z_0 may easily be calculated. For other configurations, one may make a large-scale model of the transmission line and measure Z_0 against tuning-element position using a time-domain reflectometer.

It is illuminating to take figures from Ivanek *et al.* and put them into eqn. 1. We would expect Z_0 to lie within the range 30–100Ω. Then, for a waveguide height of 0·1 in, we find that it would be possible to tune the circuit using devices with C_d between 0·3 and 1·0 pF to a frequency of 9·5 GHz. If we assume that $C_d = 0{\cdot}5$ pF, a variation of Z_0 from 54Ω to 73·5Ω will give the tuning range 10–8·7 GHz obtained.[1] This is all consistent with the results of Ivanek *et al.*[1] The range of Z_0 from 30 to 100Ω would give the wide tuning range 12·8–7·5 GHz (assuming, of course, that the other conditions for oscillation such as loading were maintained). This wider range of Z_0 (and more) could be obtained by suitably shaping the sliding short circuits,[3,4] and ensuring that they could come close to the device support post.

Conclusions: By considering the action of these circuits in terms of TEM-line resonators, a simple design technique can be established. The fact that the circuits use a waveguide below cutoff is relevant only to ensure that modes other than TEM do not occur in the resonators. It should be possible to obtain a large increase in the mechanical tuning range by a simple modification to the circuit.

M. K. McPHUN

School of Engineering Science
University of Warwick
Coventry, War., England

6th August 1969

References

1 IVANEK, F., SHYAM, M., and REDDI, V. G. K.: 'Investigation of waveguide-below-cutoff resonators for solid-state active devices', *Electron. Lett.*, 1969, **5**, pp. 214–216
2 'Transmission line', British Patent 1095103
3 McPHUN, M. K.: 'A u.h.f. tunnel diode amplifier', *Proc. IEE*, 1967, **114**, pp. 428–434
4 McPHUN, M. K.: 'A tunable X band tunnel diode amplifier', *in* '6th international conference on microwave and optical generation and amplification', *IEE Conf. rubl.*, 27, 1966, pp. 292–296

APPENDIX 11.

Microwave amplifier.

U.K. Patent No. 1 084 666,

Complete specicication published September 1967.

1,084,666

PATENT SPECIFICATION

DRAWINGS ATTACHED

1,084,666

Inventor: MICHAEL KEITH McPHUN.

Date of Application and filing Complete Specification March 25, 1966.

No. 13291/66.

Complete Specification Published September 20, 1967.

Index at Acceptance: Class H3, T(1A2:1A5:1N1:2L:2P:2RX:2T2 N1); **H1,** W(2:11:13B1:13Y).
Int. Cl.: H 03 f 3/12 // H 01 p

Microwave amplifier.

COMPLETE SPECIFICATION

We, MULLARD LIMITED, a British Company of Abacus House, 33 Gutter Lane, London, E.C.2, do hereby declare the invention, for which we pray that a patent may be granted to us, and the method by which it is to be performed, to be particularly described in and by the following statement:—

This invention relates to a microwave amplifier employing a tunnel-diode. As is known, such amplifiers are commonly of the general arrangement illustrated in Figure 1 of the accompanying drawings.

In Figure 1 a single-port amplifier comprises an input port IP through which an incoming signal is applied through a quarter-wavelength matching section of transmission-line l_1 to a tunnel-diode D. The diode functions as a negative-impedance amplifier and supplies an amplified signal back to the input port IP; with such single-port amplifiers of course, the input-port is usually connected to one port of a circulator; the incoming signal is applied to one of the other arms of the circulator and the amplified signal appears at a third arm of the circulator. In such a tunnel-diode arrangement the bandwidth of the device is restricted by means of a second matching section l_2 incorporating a series inductor L1, and a third section l_3 which is less than a quarter-wavelength long and is terminated by a capacitor C. At the junction of the two sections l_2 and l_3 a resistor R is connected across the line and this arrangement provides a stabilizing circuit which serves to damp out response of the amplifier outside the required frequency range; if such a stabilizing arrangement were not provided the amplifier would tend to oscillate at frequencies outside the desired bandwith of the amplifier.

It is, of course, necessary to provide the tunnel-diode with a voltage bias so as to fix its working point, and the bias voltage may, for example, be applied through a transmission-line stub F incorporating a radio-frequency choke RFC; in practice the source-resistance of the bias supply is sometimes influenced by the stabilizing resistance represented in Figure 1 by resistor R.

According to the present invention a tunnel-diode amplifier comprises means for applying an appropriate bias voltage to the diode so as to permit operation at a desired point on the diode current/voltage characteristic, a stabilizer resistance across the diode to reduce oscillatory tendencies, means to render the stabilizer resistance ineffective to alter the bias source D.C. resistance, an appropriate source resistance for the said bias voltage so as to permit operation at a desired dynamic gain characteristic while being otherwise ineffective at radio frequency, and a quarter wavelength input/output access path to the diode of variable characteristic impedance, for adjusting the overal gain of the amplifier.

Suitably, the amplifier may comprise an input port which is connected through said access path to the diode, an inductive section of line connected in parallel with the access path at one end across the diode, a further short-circuited section of line connected to the other end of the inductive section, and means for applying said bias voltage to the junction of the second and third sections of line.

An embodiment of the invention will now be described with reference to Figures 2 to 10 of the accompanying diagramatic drawings in which:

Figure 2 illustrates the general characteristics of a tunnel diode.

Figure 3 illustrates the general form of dynamic gain characteristics of a tunnel-diode amplifier,

Figure 4 is a simplified circuit diagram illustrating an embodiment,

Figure 5 illustrates the mechanical arrangement of the embodiment,

(Price 4s. 6d.)

Figures 6, 7, 8, 9 and 10 illustrate dynamic gain characteristics of the embodiment.

Referring now to Figure 2 of the drawings, this illustrates the current/voltage characteristic of a tunnel diode and shows at E1, E2 and E3 various working points which may be selected in an amplifier employing such a diode. Now, there are three main parameters which require to be selected when designing a tunnel diode amplifier and usually one or two of these parameters will have to be selected to match the equipment with which the amplifier is being used or to meet other requirements of the complete apparatus: in such circumstances the performance of the amplifier may be less efficient or may be less satisfactory because the third parameter, which cannot be altered without altering the other two predetermined parameters, falls outside the most suitable value.

The three parameters referred to above are the gain of the amplifier, the noise of the amplifier and the dynamic range of the amplifier. Thus, for instance, for low-noise working it is usually preferable to select a working voltage such that the diode is biased to a fairly low point on the negative-impedance portion of its characteristic illustrated in Figure 2: such a point, for instance, may conveniently be indicated by E1 on Figure 2. The gain of the amplifier may, however, so vary with the bias voltage that for the required gain the diode must be biased to a point such as E2 in Figure 2. Thirdly, however, the dynamic range of signal which the amplifier will accept may be found to require a bias point, such as E3 in Figure 2, which is different from the bias points for either maximum gain or minimum noise, and probably would be different from the bias point for the required gain which of course need not necessarily be the maximum gain achievable. As mentioned above, one of these three parameters is usually of most importance and some compromise, therefore, has to be made between the most important parameter or parameters and those of lesser importance, and it has hitherto proved impossible to produce a device in which the optimum value for each of these three parameters can be achieved simultaneously.

Referring now to Figure 3, this illustrates the general form of the dynamic gain characteristic of a tunnel-diode amplifier. As can be seen, up to a dynamic range indicated at F the gain is substantially constant but beyond that point, that is to say with input signals of a magnitude greater than that corresponding to the magnitude F, the gain at bias voltage E2 falls off rapidly due to the decreasing slope of the current-voltage characteristic of the diode. With a bias voltage at point E3, somewhat more than at E2 the gain can in fact show a slight increase before falling off due to the increasing slope of the current/voltage characteristic of the diode as the input signal swings to the left of point E3 in Figure 2. If the working point is moved to E1, Figure 2, the "hump" in the gain curve becomes more marked as illustrated by the curve E1 in Figure 3. It can thus be seen that alteration in the working point of the diode has a very marked effect upon the dynamic range of the amplifier and this range can in fact be extended by moving the working point further up the current/voltage characteristic, at the cost of some increase in gain produced by the humps illustrated in curves E1 and E3 in Figure 3.

Referring now to Figure 4, this illustrates the equivalent circuit diagram of a tunnel-diode ampilfier in coaxial line, of which the mechanical construction is illustrated in Figure 5. From an input port IP a quarter-wavelength section l_1 of transmission line leads to a diode D. Across the diode D is also connected a distributed circuit comprising a section of transmission line l_2 including a series inductor L1, and a section l_3 including a series capacitor C3. In this embodiment the section l_2 has a central conductor of reduced cross-section and also is less than one quarter-wavelength long so that the section l_2 itself forms the inductor L1.

At the junction J of l_2 and l_3 is a stub having a characteristic impedance of 20 ohms and terminated in a non-reflective load formed by a resistor R1 also of 20 ohms. Across the resistor R1 is applied the bias voltage E_b from a source not shown. This bias-voltage source is of variable resistance; such a variable resistance may be provided by an adjustable resistor connected either in series with the source, or as illustrated by a resistor R3 connected in parallel with the source across the bias supply lines. It can thus be seen that the stabilising resistor R, Figure 1, which is normally provided in a tunnel-diode amplifier is formed by R1 and is quite separate from the resistor which forms, at least in part, the source resistance of the bias supply. This, therefore, permits variation in the bias supply source resistance without affecting the stabilising resistor R1 and if the source resistance adjuster is in the form of an adjustable resistor R3 across the bias supply lines, then this also permits the bias source resistance and the bias voltage to be adjusted independently of each other. We thus have provided the facility of independent adjustment of these two parameters so as to achieve the facility of choosing our dynamic gain characteristic by adjustment of either parameter. This, of course, is par-

ticularly valuable when one of the parameters is fixed by other considerations. Thus, for instance, as described earlier, the bias voltage may have to be determined by the minimum noise permissible in the amplifier but with an arrangement according to the invention the required dynamic gain profile can nevertheless substantially be achieved by suitable adjustment of the source resistance of the bias supply.

In order that the D.C. resistance of the bias source shall not be affected by the terminating stabilizing resistor R1 of the bias stub, resistor R1 has in series therewith a capacitor C2 whilst the bias line includes a series inductor L2 so that R1 is effective only at radio frequency and not at zero frequency whilst R3 is effective only at zero frequency and not at radio frequency.

The length of the transmission line section l_2 between the diode D and the bias-supply stub is such that the impedance at the junction J appears as a short-circuit when the stabilising circuit is correctly tuned. This tuning can be adjusted by means of capacitor C3 and it will be noticed that the position of this capacitor may be different from that of the capacitor C in the conventional arrangement illustrated in Figure 1.

The diode tuned circuit itself comprises the diode capacitance C*d* which may if desired be augmented by an additional variable capacitor C*v*, the inductor L1 and the section l_2 of transmission line which extends betwen the diode and the junction J and which includes the inductor L1.

Referring now to Figure 5 this shows the mechanical construction of a tunnel-diode amplifier of which the equivalent circuit diagram has already been discussed with reference to Figure 4. In Figure 5 the input port to the amplifier is formed by the inner conductor of a coaxial line. this conductor extending into a waveguide G so as to form a probe which couples energy from the waveguide to the coaxial line. Suitably the probe P is surrounded where it projects into the waveguide G by dielectric material H so as to reduce the missmatch at the coupling probe. The section of coaxial line leading from the waveguide to the diode corresponds to the quarter-wavelength section l_1 of Figure 4. The diode D is connected between the end of the centre conductor of this section l_1 and the enclosing casing which over the section l_1 also forms the outer conductor of the coaxial line. An adjustable capacitance probe of conventional form provides the diode-circuit tuning capacitor C*v*. From the junction of the diode with this centre conductor extends a short portion of conductor which as mentioned above is in this embodiment of reduced cross-section and is shorter than one quarter-wavelength so as to have a higher inductance than the remainder of the transmission lines in this device and thereby to provide the inductor L1 of Figure 4. This central conductor of reduced cross-section leads to a further conductor K which forms the central conduction of the bias stub and which has disposed between its central conductor K and the other conductor, a tapered lossy termination which forms the resistor R1 and capacitor C2. Such a tapered load, may as is well known, be provided by a dielectric material loaded with a "lossy" material such for instance as carbonyl iron. The bias source which incorporates the variable resistor R3 of Figure 4 can be connected to the terminals E_b which correspond to the similarly-indicated terminals on Figure 1. In practice it has been found that R3 can be a normal carbon resistor, the wire "ends" or leads of which have sufficient self-inductance to constitute the inductor L2.

The portion l_3 of the transmission line section is provided by a length of coaxial line which is terminated by a conventional short circuit piston S and this section l_3 is connected to the central conductor of the rest of the system through a capacitance C3 formed between a small disc T connected at its end and the central conductor K. The value of this capacitance C3 is adjustable independently of the adjustment of the piston S so as to be able to vary the bandwidth of the amplifier without appreciably altering the resonant frequency of section l_3.

The arrangements described thus far have been concerned with the provision of means for independent variation of the bias source resistance and the bias voltage. The invention however also provides an arrangement for varying the gain of the amplifier. This is achieved by providing a section l_1 of variable characteristic impedance. As illustrated in Figure 5 one wall of the coaxial section l_1 has disposed therein a suitably-shaped member U which can be adjusted into or out of the section l_1 in the direction indicated by the arrows. This adjustment has no influence on the effective length of the section l_3 but it alters the characteristic impedance of this section very considerably. Thus, by means of variation of this characteristic impedance, the matching between the amplifier and the waveguide G can be varied and hence the overall gain of the amplifier can be varied.

It can thus be seen that the invention can provide separate independent adjustment of the gain, the working voltage point of the diode, and the dynamic range of the amplifier.

A general description will now be given of the method of setting up an amplifier such as has been described with reference

to Figures 4 and 5.

First, the diode D is removed and the stabilizing circuit is set up by adjusting C3, I_3, so as to give unity reflection at the resonant working frequency of the amplifier.

The diode is then inserted and the bias voltage E_b is adjusted to a value which provides minimum noise. This value will normally be known as it will be one of the published characteristics of whatever particular diode is in use.

The diode-circuit tuning capacitor Cv is then adjusted to give maximum gain at the working frequency.

The characteristic impedance of the transmission-line section I_1 is then adjusted, by adjusting the position of the member U, so as to give the required gain. Finally, resistor R3 is adjusted to give the required dynamic gain characteristic.

Referring now to Figure 6 this illustrates results obtained from a X-band tunnel diode amplifier constructed as has been described with reference to Figures 4 and 5; as can be seen these dynamic gain curves follow the general form illustrated in Figure 3 both the input level and the gain being shown in decibels referred to arbitrary level. For a permissible gain variation of the order of $\pm$ 1dB the largest dynamic range can be seen to be achievable with a bias voltage of 185mV but even if the gain can be allowed to vary within the limits of $\pm$ 2dB the input level, that is to say the dynamic range, is not significantly increased because the "fall-off" slope of the 190 mV curve is considerably steeper than that of the 185mV curve. The curves shown in Figure 6 will, therefore, serve to show how important is the selection of the optimum dynamic-gain working point.

Figure 7 illustrates curves, similar to those shown in Figure 6, for three voltages but it will be observed that these curves show much larger humps in their characteristics and it is, therefore, necessary to explain a further feature of these two sets of curves.

Investigation on the general form of the tunnel-diode amplifier has revealed that this type of amplifier is susceptible not to the D.C. resistance of the bias source itself in Figs. 4 and 5 effectively adjustable by R3. This is illustrated by the two sets of curves of Figures 6 and 7, Figure 6 being based on measurements taken with a bias source having a resistance of 22 ohms, whilst Figure 7 shows results obtained with the same apparatus but having a bias source resistance of 82 ohms. Comparison of the curves of Figure 7 with the curves pertaining to the same voltages in Figure 6 show how the increase in the D.C. resistance of the bias source has increased the amplitude of the gain hump and the similarity of the effect on this gain hump of increasing the bias voltage to the effect of increasing the bias source resiistance will be apparent.

This is illustrated in a slightly different fashion in Figures 8 and 9 which show curves obtained with a constant bias voltage of 175mV and 180mV respectively but with four different bias source resistances of 10, 22, 47 and 82 ohms.

The similarity between the shapes of the curves of constant bias source resistance and different bias voltages as shown in Figures 6 and 7 and the curves for constant bias voltage and different bias source resistance as shown in Figures 8 and 9, is illustrated in Figure 10 which shows curves for 170mV at 82 ohms and 185mV at 22 ohms.

WHAT WE CLAIM IS:—

1. A tunnel-diode amplifier comprising means for applying an appropriate bias voltage to the diode so as to permit operation at a desired point on the diode current/voltage characteristic, a stabilizer resistance across the diode to reduce oscillatory tendencies, means to render the stabilizer resistance ineffective to alter the bias source D.C. resistance, an appropriate source resistance for the said bias voltage so as to permit operation at a desired dynamic gain characteristic while being ineffective otherwise at radio frequency, and a quarter wavelength input/output access path to the diode of variable characteristic impedance for adjusting the overall gain of the amplifier.

2. An amplifier as claimed in either preceding claim, comprising an input port which is connected through said access path to the diode, an inductive section of line connected in parallel with the access path across the diode, a further, short circuited section of line connected at the other end of the inductive section, and means for applying said bias voltage to the junction of the second and third sections of line.

3. An amplifier as claimed in Claim 2 wherein the said stabilizer resistance is also connected across the junction for radio frequenncies, but not at zero frequency.

4. An amplifier as claimed in Claim 3 wherein the said further section of line includes an adjustable series capacitance.

5. An amplifier as claimed in Claim 3 or Claim 4 wherein the means for applying the bias voltage comprise a section of transmission line in the form of a stub which is terminated in the stabilizing resistance which is equal to the characteristic impedance of the line forming the stub.

6. An amplifier as claimed in Claim 5 wherein the last mentioned section includes an adjustable resistor to shunt the source, and a series inductor, the terminating resistance offering an open circuit to direct cur-

rent by virtue of having a series capacitor with it.

7. An amplifier as claimed in any of Claims 1 to 7 wherein the said access path is a coaxial transmission line with a positionably adjustable member to produce variation of its characteristic impedance which varies the overall gain of the amplifier.

8. An amplifier according to claim 4 wherein the adjustable capacitance comprises a movable capacitance disc and the short circuit position in said further line section is independently adjustable.

9. A tunnel diode amplifier substantially as described with reference to Fig. 5 of the accompanying drawings.

10. An amplifier as claimed in Claim 1 constructed and arranged to operate substantially as herein described with reference to Figures 4 to 10 of the accompanying drawings.

G. V. CARCASSON,
Chartered Patent Agent,
Mullard House,
Torrington Place,
London, W.C.1.
Agent for the Applicants.

Sheerness: Printed for Her Majesty's Stationery Office by Smiths Printers and Duplicators.—1967.
Published at the Patent Office, 25 Southampton Buildings, London, W.C.2, from which copies may be obtained.

1084666 COMPLETE SPECIFICATION
5 SHEETS *This drawing is a reproduction of the Original on a reduced scale*
Sheet 1

FIG. 1.

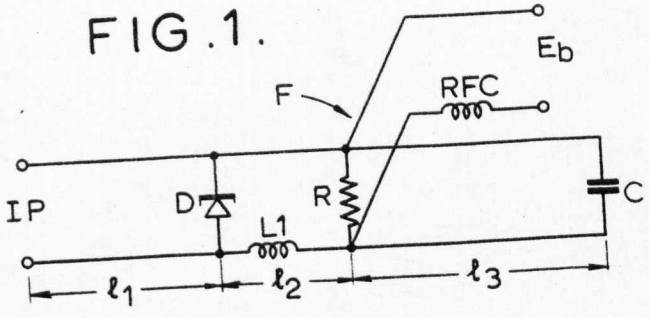


FIG. 2.

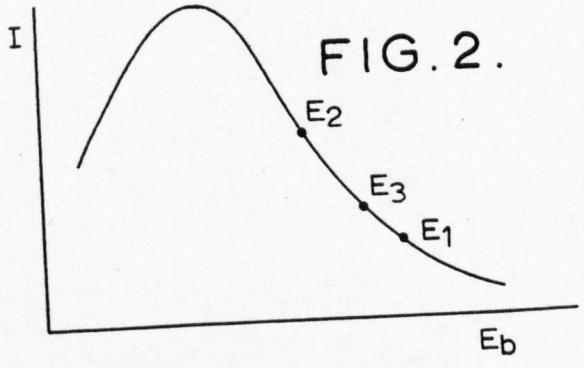
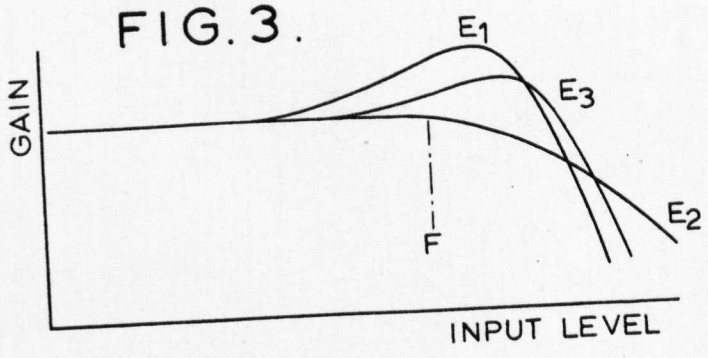


FIG. 3.



1084666 COMPLETE SPECIFICATION
5 SHEETS This drawing is a reproduction of the Original on a reduced scale
Sheets 2 & 3

FIG. 4.

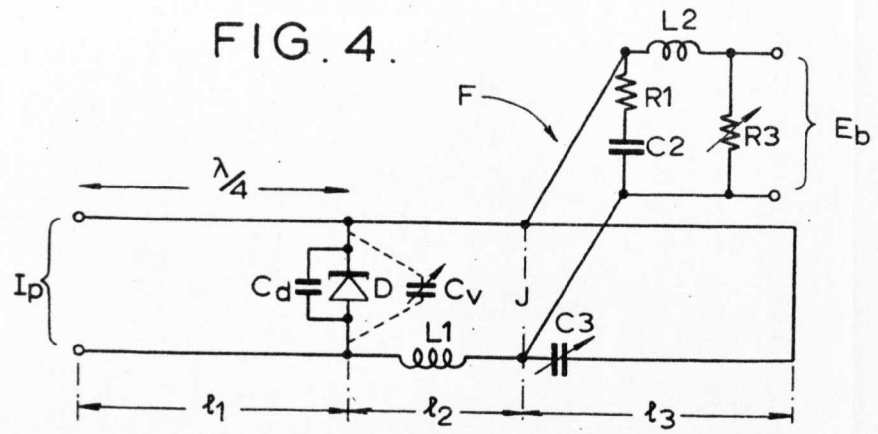


FIG. 5.

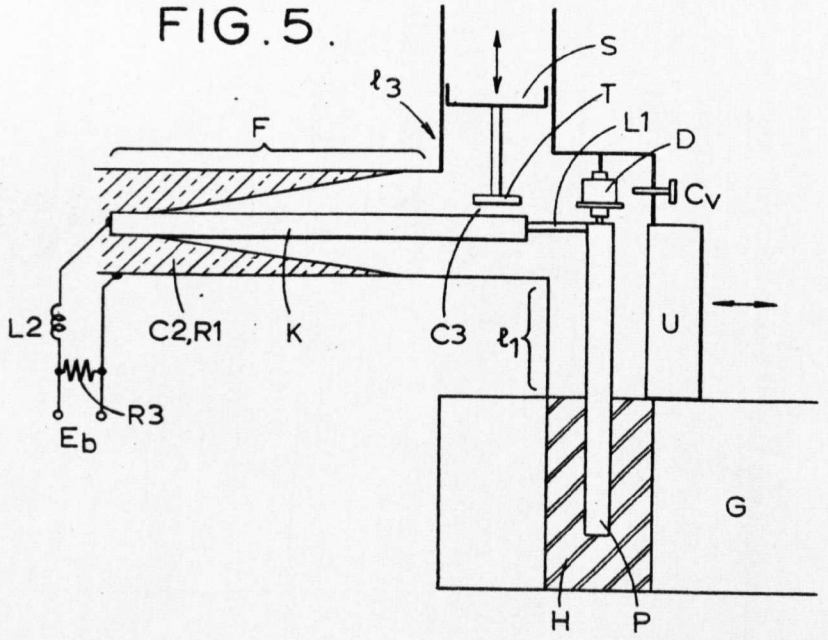


FIG. 6.

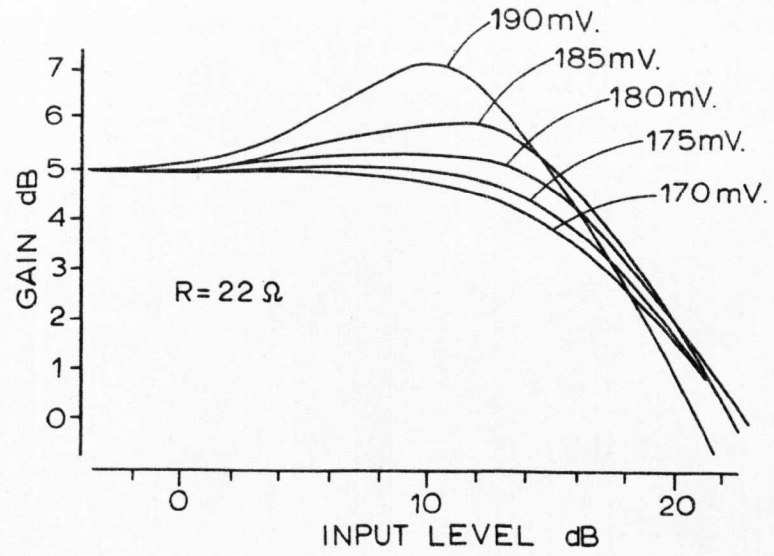
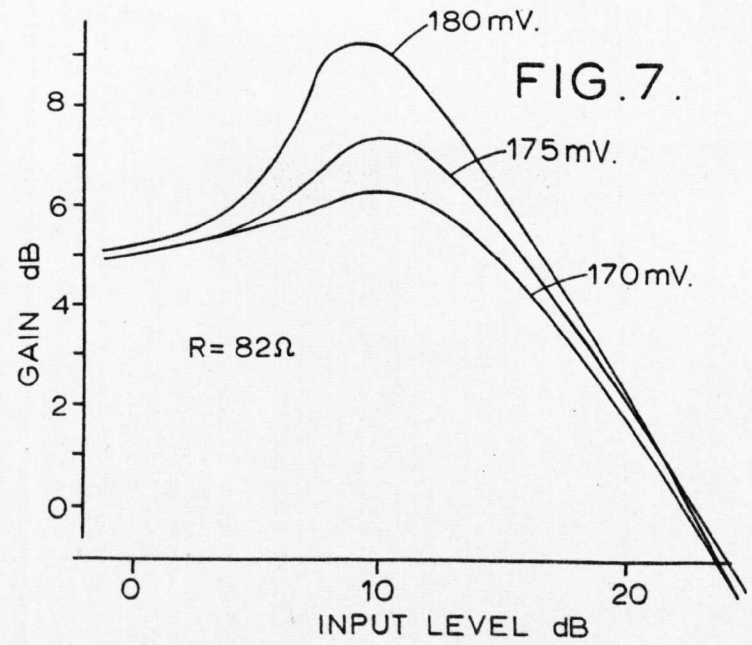


FIG. 7.



1084666 COMPLETE SPECIFICATION
5 SHEETS *This drawing is a reproduction of the Original on a reduced scale*
Sheets 4 & 5

FIG. 8.

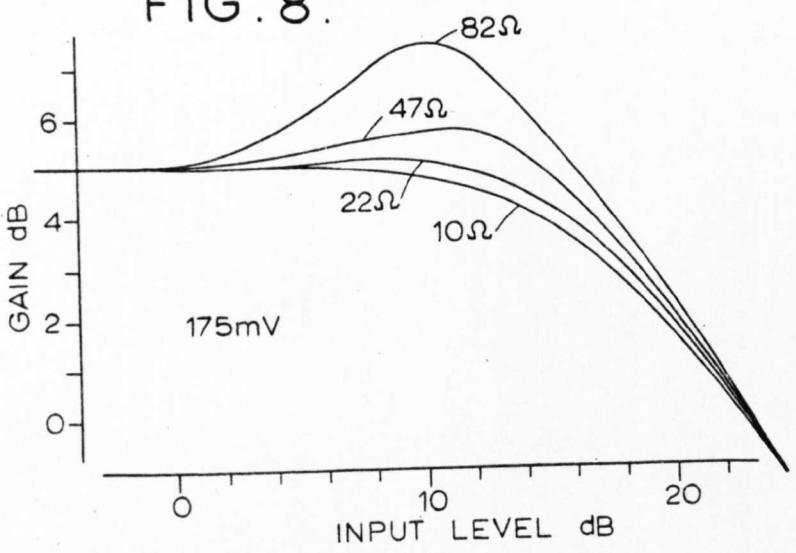


FIG. 9.

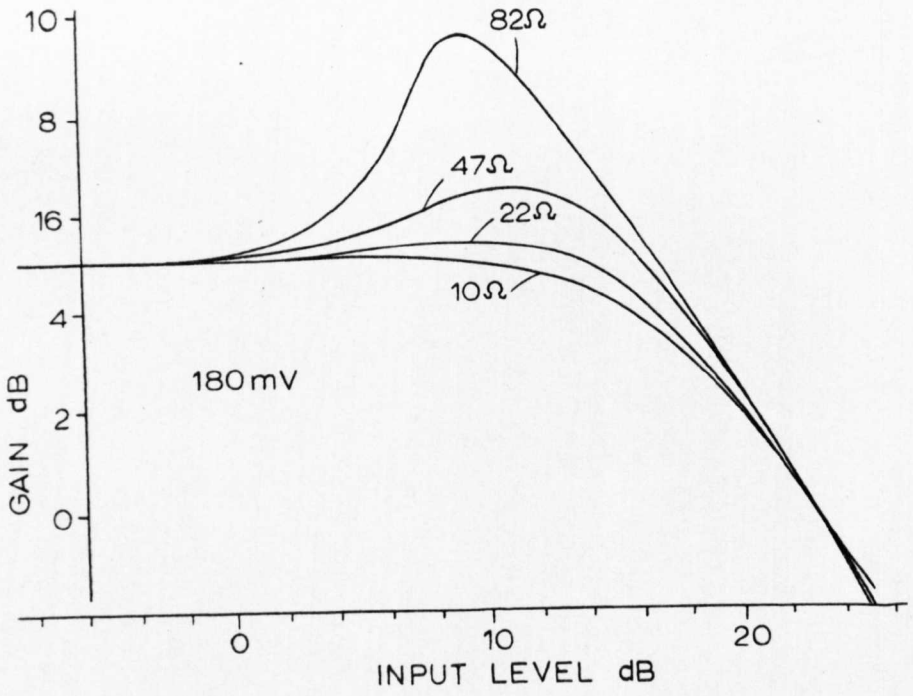
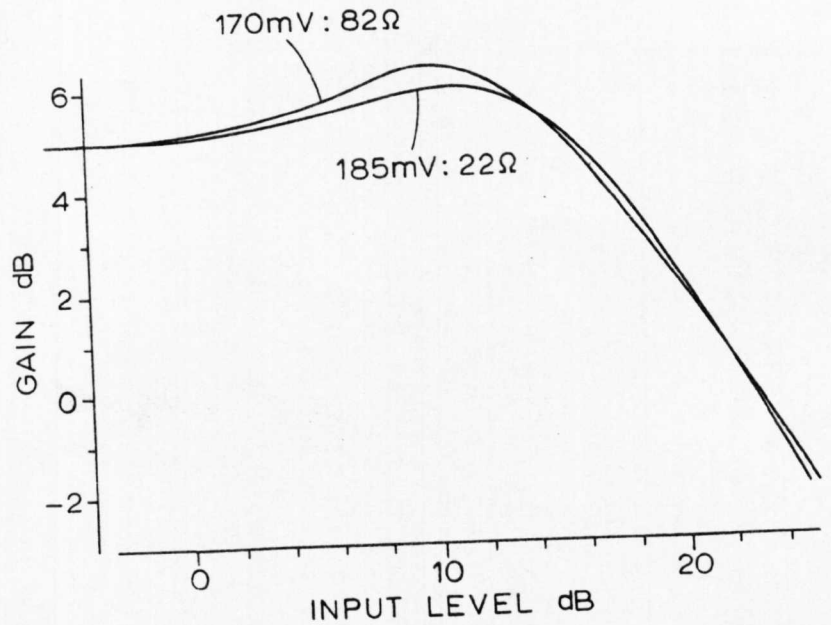


FIG. 10.



APPENDIX 12.

Variable-impedance transmission-line device.

U.K. Patent No. 1 095 103,

Complete specification published December 1967.

1,095,103

PATENT SPECIFICATION

DRAWINGS ATTACHED

Inventor: MICHAEL KEITH MCPHUN

1,095,103

Date of Application and filing Complete Specification: March 25, 1966.

No. 13292/66.

Complete Specification Published: Dec. 13, 1967.

Index at acceptance: —**H1** W(2, 7, 11, 13BX, 13Y)

Int. Cl.: —**H 01 p 5/04**

COMPLETE SPECIFICATION

Variable-impedance Transmission-line Device

We, MULLARD LIMITED, a British Company of Abacus House, 33 Gutter Lane, London, E.C.2., do hereby declare the invention, for which we pray that a patent may be granted to us, and the method by which it is to be performed, to be particularly described in and by the following statement: —

THIS INVENTION relates to a device for providing a variable-impedance section of transmission line. The invention is applicable to coaxial transmission line and also to symmetrical double-ground-plan strip-line which as is known can be regarded as a developed form of coaxial line.

In some arrangements it is of advantage to provide a quarter-wave matching section of which the characteristic impedance can be varied: such a variable-impedance section can be used, for example, in a reflection amplifier such as a tunnel-diode amplifier, where the facility of varying the impedance of a quarter-wave section connected between the tunnel-diode and the input/output port of the amplifier provides means for varying the gain of the amplifier as a whole.

It is known from the specification of British Patent 706090 (page 2, last paragraph) that the movement of various members in proximity to the conductors of an open wire line alters the characteristic impedance of a section of the line, e.g. according to a sinusoidal law, and that a single movable body can be made to alter the impedance of two lines at once.

A facility is particularly desirable, however, in wide-band devices, namely to maintain a specified VSWR over a specified bandwidth of a single run of transmission line. Hitherto, it has been necessary to provide separate elements each in the form of a variable-impedance device, and to adjust them by a "hit and miss" technique until the desired result has been achieved, and the teaching of the above specification is of no help in containing VSWR variations within reasonable limits.

The present invention provides a device which permits of rapid adjustment of successive quarter-wave matching sections so as to obtain a desired impedance transformation.

According to the present invention, a device providing a length of coaxial line or a symmetrical double-ground-plane strip-line having an adjustable characteristic impedance comprises such a transmission line at least one half-wavelength long at a frequency passed by the line, together with a single member extending along the transmission line and adjustable transversely thereof in position towards and away from the central conductor or conductors of the line, wherein the member is arranged so that as it is moved transversely into or out of the line the rate of variation of characteristic impedance is different in successive quarter-wavelength sections of the line.

The member may be of such shape as to provide in successive quarter-wavelength sections of the line, parts having different proportions so as to provide different characteristic impedances for the two sections. Alternatively, the member may be arranged to provide a tapered characteristic impedance along part or all of the length of transmission line.

The member may be movable into or out of the line along any suitable path which may be arcuate or linear.

Embodiments of the invention will now be described by way of example, with reference to the accompanying diagrammatic drawings in which: —

Figure 1 illustrates an arrangement of one type of symmetrical strip-line,

[*Price 4s. 6d.*]

Figure 2 is a further view of such a line,

Figure 3 illustrates strip-line embodiment of the invention,

Figures 4, 5, 6 and 7 illustrate various profiles of the adjustable member.

Figure 8 illustrates an embodiment incorporating a different type of symmetrical strip-line,

Figures 9 and 10 illustrate a coaxial-line embodiment,

Figures 11, 12 and 13 illustrate a further coaxial-line embodiment,

Figures 14, 15 and 16 illustrate an adjustable member and

Figure 17 illustrates a further strip-line embodiment.

Referring to Figures 1 and 2 these illustrate in schematic form what is usually known as a "tri-plate" line and in the form shown in these two Figures the line is of the type wherein two metal conductor strips A are supported on a strip B of dielectric which is in turn held centrally, by means not shown, between two ground-plane strips G.

As shown in Figure 3 an embodiment of the invention incorporating the construction of stripline illustrated in Figures 1 and 2 again comprises two ground-planes G, a central dielectric strip B and two conductor strips A of which only one appears on Figure 3. Between the two plates G is disposed a member C which is in two parts, having the profiles indicated at C1 on Figure 4 and C2 on Figure 5, these latter two figures being cross-sectional views taken respectively on the lines IV—IV and V—V of Figure 3. As can be seen from Figures 4 and 5 the portions C1 and C2 of the adjustable member C have different profiles at their location adjacent the conductors A and as the member C is moved into or out of the device along a linear path in a direction indicated by the arrows in Figures 4 and 5 not only will the characteristic impedance of the line be varied but the variation in the line section which includes the part C1 will be different from the variation in the other line section which includes the part C2. It can thus be seen that adjustment of the single member C affords the simultaneous adjustment, over the two sections of line, of characteristic impedance according to two different laws. In this text the term "law" is understood to mean the dependence of the characteristic impedance upon the distance by which the member is adjusted into or out of the transmission line section. As can be seen from Figures 3, 4 and 5, the central dielectric strip B is cut away adjacent the member C so as to enable this member to approach or to embrace the conductor strips A.

Figures 6 and 7 illustrate two further profiles of parts C3 and C4 of an adjustable member C which are also applicable to the form of strip-line illustrated in Figures 1 and 2.

Figure 8 illustrates a modification where the member C, which has a portion C3 having the same configuration as the member C3 in Figure 6, is adjustable in a system comprising a single centre conductor A supported by conventional means not shown.

Figures 9 and 10 illustrate an embodiment for use with a coaxial line system. In this arrangement the outer conductor of a coaxial transmission line is formed in a metal block E and comprises a semi-cylindrical wall F and two parallel walls H, thus forming a trough having a cylindrical end wall F. In Figure 9, a part C5 is adjustable into and out of this trough-like cavity along a linear path in a direction of the arrows so as to approach or recede from the central conductor J of the coaxial line: this central conductor is of course supported concentrically with the wall F and with the remainder of the line by adjustable dielectric supports not shown.

Figure 10 illustrates a second profile in which the part C6 is spaced further from the central conductor by providing it with a part-cylindrical surface K of greater radius than the corresponding part-cylindrical surface K on the member C5 in Figure 9. The arrangement of such a coaxial device would be in general similar to that illustrated for the strip-line device in Figure 3, the members C5 and C6 form in fact portions of a single adjustable member in substantially the same manner as the portions C1 and C2 form parts of a single member C in Figure 3.

Figures 11, 12 and 13 illustrate a further embodiment in coaxial-line arrangement in which a metal block E again provides the outer conductor of the line. Figures 11 and 12 illustrate a part C7 of an adjustable member which is pivoted on a pivot P extending parallel to the length of the line so as to be movable into and out of the line, towards or away from the central conductor J of the line along an arcuate path about the pivot P: these two Figures illustrate two positions of the part C7. The part C7 is associated with a part C8 of the adjustable member, see Figure 13, having at its location adjacent the central conductor J a profile which differs from that of the part C7.

In the arrangements illustrated in Figures 4 to 7 and 9 to 13 the member C has parts which are of different profile adjacent the inner part of the line, but this feature is not essential and for some profiles the requirement of different impedance variations in two or more sections of the line can be met by providing parts of the member C which are of the same profile but which have different spacings relative to the central con-

ductor or conductors. Such an arrangement is illustrated in Figures 14, 15 and 16 where a member C has parts C9 and C10 of the same general profile as the part C3 in Figures 6 and 8. The parts C9 and C10 have identical profiles at their inner sides but the proportions of the parts differ so that the part C9 will in use project further into the transmission line, not shown, than will the part C10. The different proportions of these two parts is emphasized by the two cross-sections illustrated in Figures 15 and 16.

Usually, the successive sections of line will be quarter-wavelength sections but it will be apparent that the invention is applicable to adjustable-impedance devices where at least one of the sections is a multiple of a quarter-wavelength long. Although for ease of description only devices comprising two successive sections have been described and illustrated, three or more sections may be arranged together so as all to be simultaneously adjustable by means of a common member C having the appropriate member of differently-proportioned parts. Further, if desired the profile of any particular part of the member C may vary along the line from one end of a section to the other instead of being of constant profile throughout its length.

If desired, instead of being formed into successive quarter-wavelength parts the member C may vary continuously from one end to the other so as to provide an adjustable "tapered" line of which the impedance tapers continuously from one end to the other.

Figure 17 illustrates a further embodiment in which a member C is movable into and out of a stripline having ground-planes G and a single central conductor A. The member C has the same cross-section throughout its length and is pivoted at one end of the device on a shaft P, which extends perpendicularly to the longitudinal axis of the line, so it can be adjusted into and out of the line along an arcuate path indicated by arrows on Figure 17. Such an arrangement provides along the length of the transmission line a tapered, that is a continuously-varying characteristic impedance, except of course when the member C is in a position when it is parallel with the line.

It will be observed that, the parts of the member C in the Figures are shown as spaced from the ground-planes or from the outer coaxial conductor: in practice such spacing is quite permissible because in TEM modes such as these the current flow is in an axial direction in the inner and outer conductors and does not cross the gap between the member C and the outer or ground-plane conductor.

It will be appreciated that although various embodiments, varying in profile and method of movement of the member C, have been described, these embodiments serve to illustrate features which may be incorporated in different combinations from the particular combinations illustrated in the Figures. Thus, for instance, the tapered-impedance facility provided by the arrangement of Figure 17 could equally well be arranged for linear movement in and out of the transmission line instead of arcuate movement: the whole member may thus vary continuously along its length either in profile or in spacing from the central conductor or in both so as to provide a tapered impedance and this tapered impedance itself may be variable by either linear or arcuate movement of the member.

Because of the basic similarity between coaxial line and strip line many of the features described above are applicable to both kinds of line; this is apparent, for instance, if Figures 9 and 10 are compared with Figures 5 to 7. Again, it is obvious that the arcuate movement of Figure 17 could be applied to coaxial line or that the arcuate movement of Figures 11, 12 and 13 could be applied to strip-line. Obviously, some features of the invention are inherently more suitable for one form of line than another, and the choice of feature will also depend upon the parameters of the apparatus with which the device is to be used; no attempt therefore has been made to illustrate or describe embodiments examplifying all the combinations or permutations of the various features of the invention.

WHAT WE CLAIM IS:—

1. A device for providing a length of coaxial line or a symmetrical double-ground-plane strip-line having an adjustable characteristic impedance, comprising such a transmission line at least one half-wavelength long at a frequency passed by the line, together with a single member extending along the transmission line and adjustable in position transversely thereof towards and away from the central conductor or conductors of the line, wherein the member is arranged so that as it is moved transversely into or out of the line the rate of variation of characteristic impedance is different in successive quarter-wavelength sections of the line.

2. A device as claimed in Claim 1 wherein the member has two parts which are arranged in successive quarter-wavelength sections of the line and which have different proportions so as to provide different characteristic impedances for the two sections of line.

3. A device as claimed in Claim 2 wherein the parts of the member are each of uniform cross-section so as to provide a uniform characteristic impedance along each section of line.

4. A device as claimed in Claim 2 wherein each part of the member is non-uniform so as to provide a tapered characteristic im-

pedance along the associated section of line.

5. A device as claimed in Claim 1 wherein the member is arranged to provide a tapered characteristic impedance along part or all of the length of transmission line.

6. A device as claimed in any preceding Claim wherein the member is adjustable into or out of the line along a linear path.

7. A device as claimed in any of Claims 1 to 5 wherein the member is adjustable into or out of the line along an arcuate path.

8. A device as claimed in Claim 7 wherein the member is pivoted about an axis extending parallel to the longitudinal axis of the line.

9. A device as claimed in Claim 7 wherein the member is pivoted about an axis extending perpendicular to the longitudinal axis of the line.

10. A device as claimed in Claim 1 constructed and arranged to operate substantially as herein described with reference to any of the accompanying drawings.

G. V. CARCASSON,
Chartered Patent Agent,
Mullard House,
Torrington Place,
London, W.C.1.
Agent for the Applicants.

Leamington Spa: Printed for Her Majesty's Stationery Office by the Courier Press.—1967.
Published at The Patent Office, 25, Southampton Buildings, London, W.C.2, from which copies may be obtained.

1095103 COMPLETE SPECIFICATION
4 SHEETS *This drawing is a reproduction of the Original on a reduced scale*
Sheets 1 & 2

FIG.1.

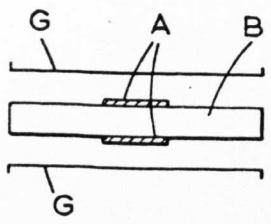


FIG.2.

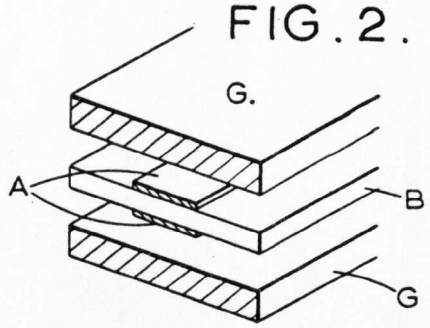


FIG.3.

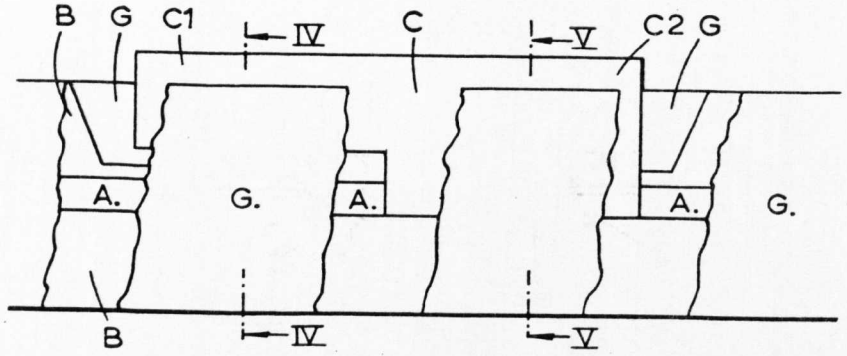


FIG.4.

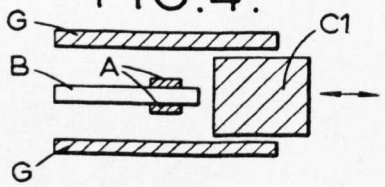


FIG.5.

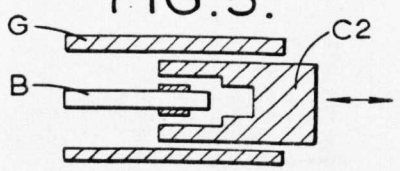


FIG.6.

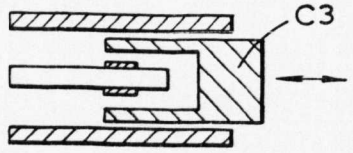


FIG.7.

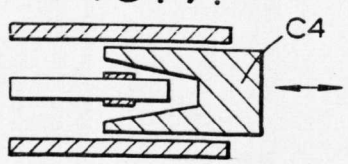


FIG.8.

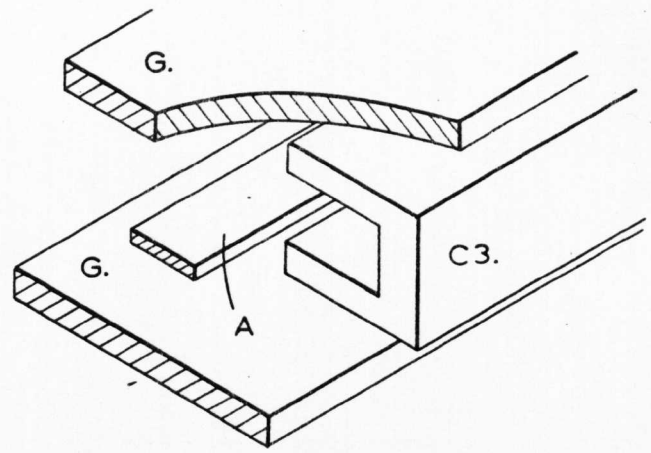


FIG.9.

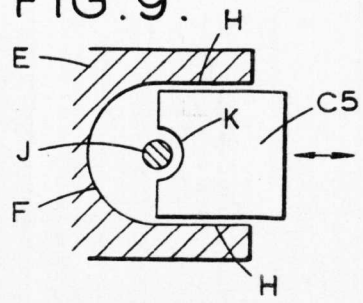


FIG.10.

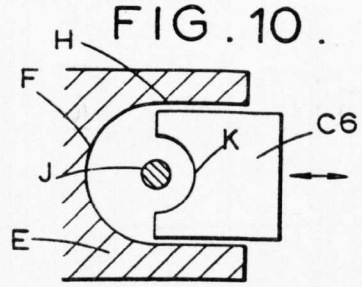


FIG.11.

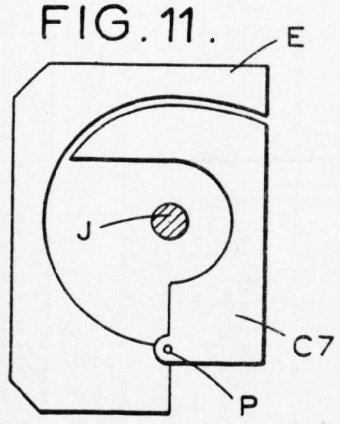
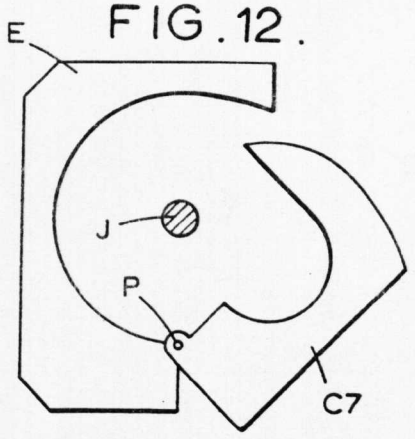


FIG.12.



1095103 COMPLETE SPECIFICATION
4 SHEETS This drawing is a reproduction of the Original on a reduced scale
Sheets 3 & 4

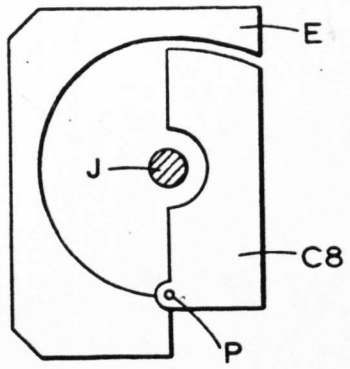


FIG. 13.

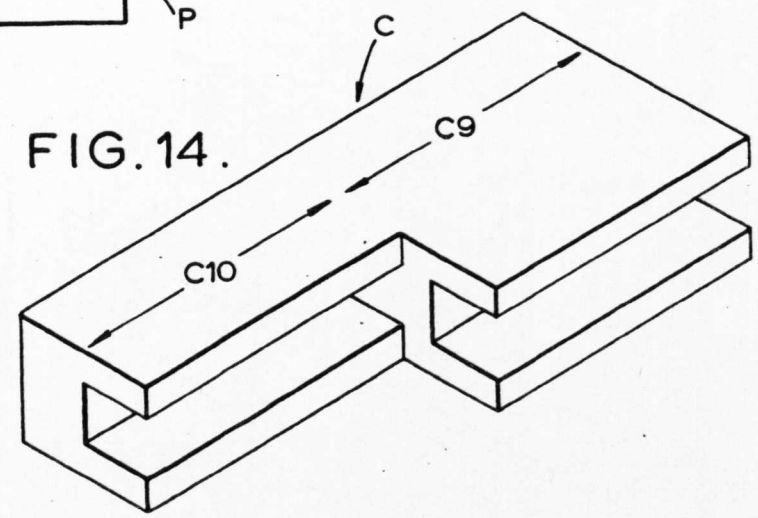


FIG. 14.

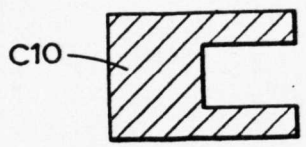


FIG. 15.

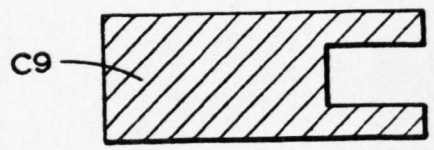
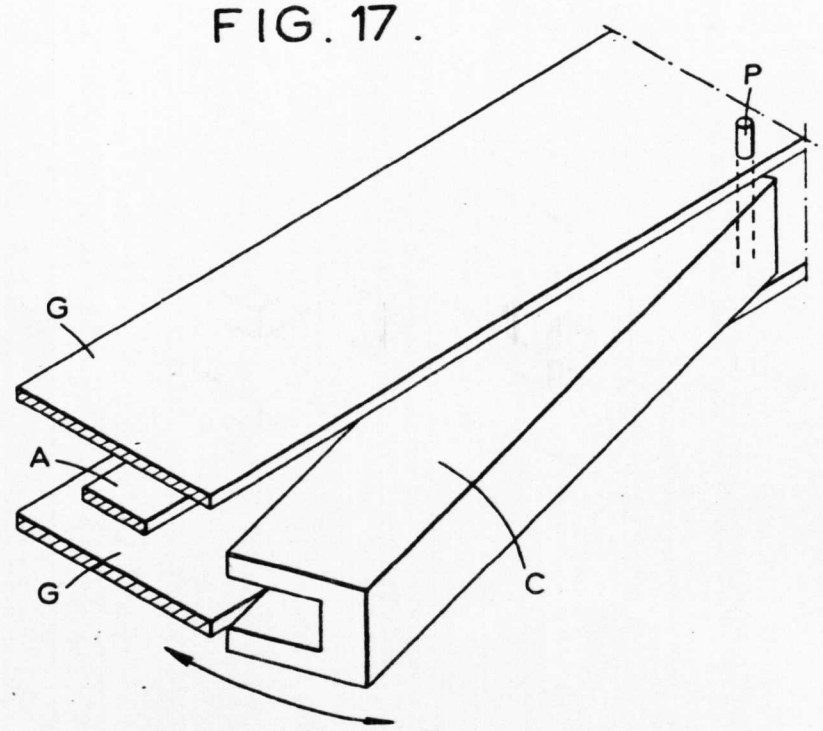


FIG. 16.

FIG. 17.



APPENDIX 13.

Tunnel diode amplifiers.

Chapter 7 of the book "Low noise microwave amplifiers" by Daglish et. al., a monograph published for the I.E.E. by Cambridge University Press, 1968.

7 TUNNEL-DIODE AMPLIFIERS

7.1 Introduction

In 1957, Esaki (1958) discovered that quantum-mechanical tunnelling in a degenerate-semiconductor p–n junction could provide a negative-resistance characteristic. The tunnel diode was then rapidly developed, primarily for switching applications, and soon reached a high degree of sophistication. However, it has not been adopted for the new generation of computers and in practice the major use is in relatively low-noise microwave amplifiers. In this application it has considerable advantages, being of a very simple construction and offering reasonable noise factors of 3–6 dB combined with good bandwidth. The early limitations of the tunnel-diode amplifier (t.d.a.) were its low power-handling capability and a tendency to fragility unless a good device design was achieved, but both these aspects have improved steadily during the past few years.

The tunnel diode has the voltage-controlled negative-resistance characteristic shown in Fig. 30, in which the main parameter is the peak current. The most useful property of the diode is that the negative-resistance part of the characteristic results from quantum-mechanical tunnelling, and is independent of frequency throughout the microwave spectrum.

In the tunnelling process, electrons can penetrate the high-electric-field region associated with the p–n junction barrier. This can only take place in degenerate or very heavily doped semiconductors in which the bottom of the conduction band on the n type side of the junction is lower than the top of the valence band on the p type side. To increase the tunnelling probability it is also necessary to have a sharp discontinuity between the n and p type materials in order to reduce the width of the transition region to 50–100 Å. With a slight forward bias applied to such a diode, there is an overlap between the relevant conduction and valence bands and a large tunnelling current flows. As the

bias is increased, there is less overlap and the current decreases until the normal forward characteristic of a p–n junction is resumed.

When suitably biased in the negative-resistance region, the tunnel diode has the equivalent circuit of Fig. 31 *a* (Hall, 1960). The junction of the tunnel diode is represented by a negative resistance $-R_n$ (or

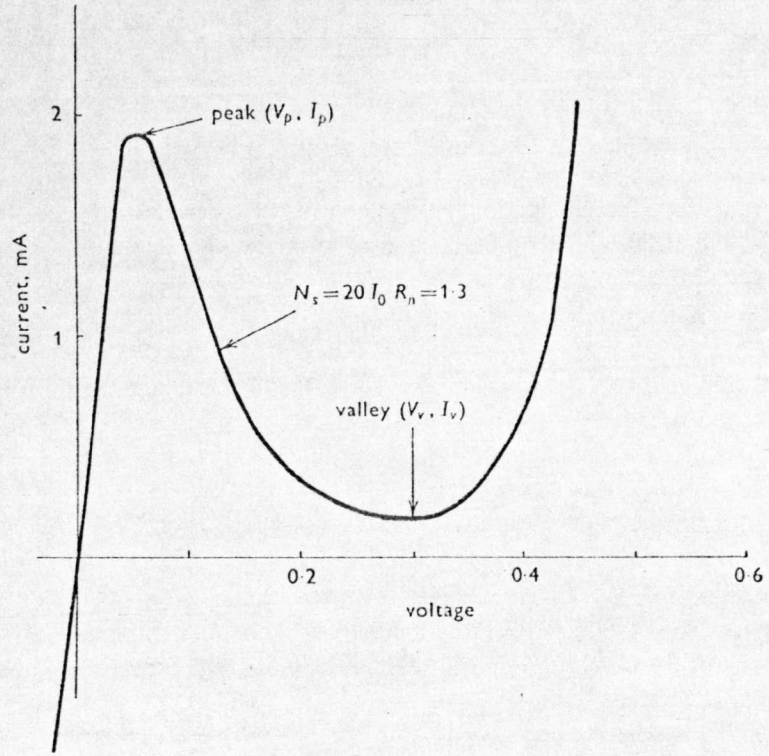


Fig. 30 Typical voltage/current characteristics of germanium microwave tunnel diode (Mullard type AEY 15)

conductance $-G_n$) in parallel with the junction capacitance C_j, where R_n is a positive number. R_s and L represent the unavoidable resistance and inductance in series with the diode.

The diode impedance $Z(\omega)$ in Fig. 31 *a* is given by

$$Z(\omega) = R_s - \frac{G_n}{G_n^2 + \omega^2 C_j^2} + j\left(\omega L - \frac{C_j}{G_n^2 + \omega^2 C_j^2}\right) \qquad (7.1)$$

The diode can be used as a negative-resistance amplifier at any frequency below the resistive cutoff frequency f_{r0}, at which point $\mathrm{Re}\{Z(\omega)\} = 0$. The series, or self-resonant, frequency f_{x0} gives $\mathrm{Im}\{Z(\omega)\} = 0$. Provided that

$$L < C_j R_n R_s \qquad (7.2)$$

$$f_{x0} > f_{r0} \qquad (7.3)$$

and it will be relatively easy to design a stable amplifier using the diode. Under these conditions, the simple equivalent circuit of Fig. 31 *b* is a close approximation for frequencies from zero to $f_{r0}/3$, and R and

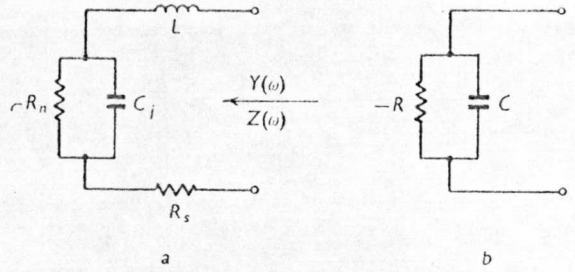


Fig. 31 Equivalent circuits of tunnel diode

a Equivalent circuit
b Approximate equivalent circuit, valid for $f < f_{r0}/3$

C differ little from R_n and C_j (McPhun, 1967). This simple equivalent circuit of constant negative resistance in parallel with constant capacitance can be used as a sound basis for initial amplifier design, but the full equivalent circuit must be used when considering out-of-band performance, which affects amplifier stability (Scanlan, 1966).

7.2 Construction of tunnel diodes

At present most diodes are fabricated by alloyed-junction techniques, in which the junction is formed by alloying a small n-doped bead into the surface of a p-type semiconductor wafer. The wafer is soldered into the housing and the connection soldered to the bead. A bead of about 25 μm diameter is used in order to keep the junction capacitance low, the capacitance being directly proportional to the wetted area of the

alloyed bead. The junction capacitance is further reduced by electrolytic etching of the junction, which also allows the peak current to be reduced to the desired value.

Etching to achieve much less than 1/20 of the initial peak-point current clearly gives a weak structure, and it is apparent that diodes having resistances above a few ohms are likely to be quite fragile. These considerations lead to the choice of a low etching ratio (ratio of initial to final peak-point current) and this results in a satisfactory mechanical and environmental life-test performance.

The housing in which the tunnel diode is mounted is of considerable importance since it contributes the major part of the series inductance and capacitance of the diode. A low series inductance results in a high self-resonant frequency f_{x0} for the diode, and cases have been designed to give a low inductance by using a thin, small-diameter, ceramic annulus. The inductance can be further reduced by using short connections of rectangular cross-section to connect the p–n junction to the housing. The use of rectangular-cross-section foil can also increase the mechanical strength of the device in its weakest direction, at right angles to the foil in the plane of the wafer surface. Other methods of reinforcing the junction by using glass encapsulation and similar techniques are also employed. These designs have resulted in series inductances of approximately 100 pH, and case capacitances of 0·25 pF. The construction of such a diode is shown in Fig. 32*a* and Plate 17.

Silicon, germanium, indium-antimonide and gallium-antimonide tunnel diodes have all been fabricated by this basic alloyed-junction method, and a wide range of devices are available. For low-noise amplifiers, gallium-antimonide diodes can ideally offer a better noise performance, but, since gallium-antimonide technology is less well advanced, most diodes are based on germanium.

It would appear that diodes based on alloying techniques are now virtually fully exploited and new technologies must be used to obtain improved performance. The fragility of the etched-mesa junctions produced in the present alloyed diodes restricts their use to X band frequencies and lower. Because the encapsulation forms an essential part of the strength of the device, these diodes are not suitable for

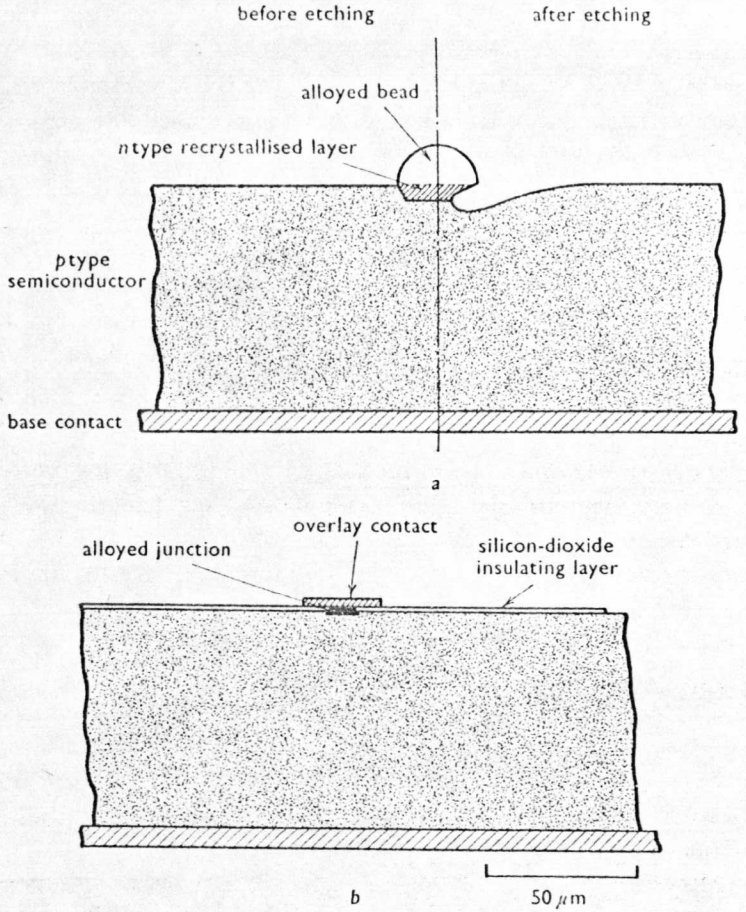


Fig. 32 Tunnel-diode construction
a Alloyed device
b Planar device

incorporation into an integrated-circuit type of amplifier, and so the true junction performance, which is degraded by the parasitics of the encapsulation, cannot be realised. However, planar technology has advanced rapidly in the past few years, and it is now possible to consider a planar tunnel diode in which the junction is defined by an oxide mask without subsequent etching or processing. It appears likely that

germanium will continue to be the most suitable material, and oxide masking and photolithography can now be used to fabricate junction areas of a few micrometres diameter. Diodes have been made by this process with f_{r0} up to 50 GHz, and, although the yield is low at present, this is an important change in technology. The structure of this type of diode is shown in Fig. 32*b*.

Devices of this type are significant in offering improved mechanical strength compared with existing mesa diodes. As planar methods improve, and possibly with the introduction of electron-beam techniques, it is becoming possible to produce smaller junctions leading to satisfactory amplifiers for frequencies above X band. These diodes are in chip form, free of encapsulation, and can be incorporated into thin-film hybrid circuits. Although a completely monolithic microwave integrated circuit cannot be formed in germanium, it is now possible to grow germanium islands in a semi-insulating gallium-arsenide substrate by means of contour epitaxy. With this isolation, and deposited transmission lines, a monolithic tunnel-diode amplifier circuit becomes feasible.

7.3 Amplifier design

The use of the circulator-coupled reflection amplifier is now universal for the t.d.a., as this has a wider bandwidth and a lower noise factor and is less dependent on load and generator impedances than a t.d.a. without a circulator.

The basic form of a circulator-coupled t.d.a. is as shown in Fig. 33*a*, where the circulator isolates the reflection amplifier from both the generator and the load. The reflection amplier shown in Fig. 33*b* terminates the transmission line of characteristic resistance R_0 in a negative resistance $-R$, giving a reflection coefficient Γ which is greater than unity:

$$\Gamma = \frac{R+R_0}{R-R_0} \qquad (7.4)$$

The power gain of the circulator-coupled amplifier is then Γ^2.

To use a tunnel diode as a reflection amplifier, the designer must do the following four things:

(*a*) bias the diode, using a load line providing a single-valued intersection in the negative-resistance region of the I/V characteristic

(*b*) present the transmission line with a suitable negative resistance over the passband

(*c*) reduce Γ to acceptable limits outside the passband

(*d*) maintain stability.

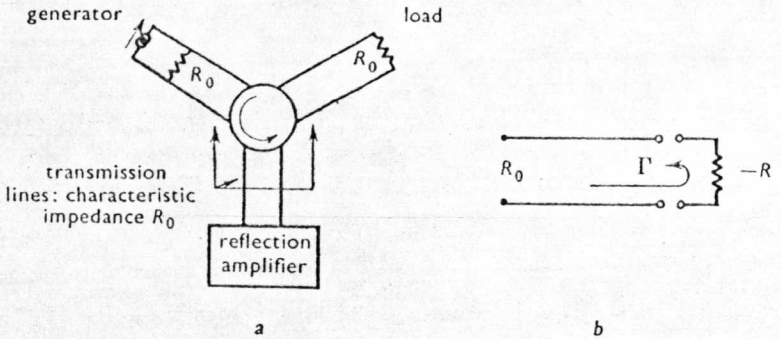


Fig. 33 Details of tunnel-diode amplifier
a Basic circulator-coupled amplifier
b Reflection amplifier

Because the tunnel diode presents a negative resistance at its terminals from d.c. to the resistive cutoff frequency, oscillations may occur anywhere within this range, so the maintenance of stability has been a major problem for the circuit designer. Oscillations may occur owing to resonance within the reflection amplifier, or owing to interactions with the circulator, generator and load. Eqn. 7.4 will only apply to a 'perfect' circulator with characteristic resistance R_0 at all frequencies. All practical circulators have a limited passband, and oscillations are most likely at the band edges.

To avoid generator and load interactions, 4- or 5-port circulators are frequently used. It is now common practice to provide a stabilising circuit either in parallel or in series with the diode. The stabilising circuit provides a resistive load for the diode at all frequencies except for the desired passbands. However, by using a sufficiently wideband circulator, it is just possible to design a satisfactory amplifier without a stabilising circuit. A basic lumped-element stabilising circuit of the

parallel type is shown in Fig. 34*a* and a transmission-line version in Fig. 34*b*. These have the general form of a bandstop filter interposed between the diode and the stabilising resistance R_{stab}.

A matching network is often required between the circulator and diode to give the required bandpass characteristics, and the complete t.d.a. has the form shown in Fig. 35. It is often convenient to include the bias circuit in with the stabilising circuit.

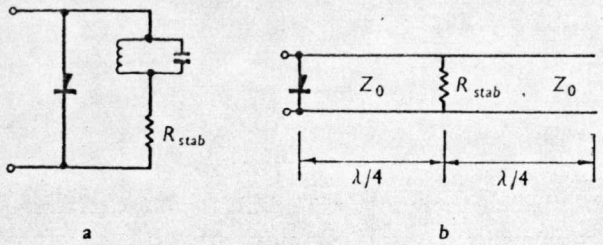


Fig. 34 Shunt-type stabilising circuits
a Lumped element
b Transmission line

It is evident that all the frequency-variable circuit elements will together determine the amplifier frequency response. However, much useful theory and design has been based on the assumption of the 'perfect' circulator (Scanlan and Lim, 1964 and 1965, and Hamasaki, 1965). This assumption is not now required, as Okean (1966*a*) has described a powerful synthesis technique which takes into account all the circuit elements including the bias circuit and the circulator.

T.D.A.s have been constructed in waveguide, coaxial and stripline, and there are many papers giving constructional details, some examples being Hamasaki (1965), McPhun (1966 and 1967), Lee (1967), Easter (1965) and Okean (1966*b*). A typical amplifier is shown in Plate 18.

7.4 Amplifier performance

Here we will review the factors affecting attainable t.d.a. performance, and examine what has been achieved so far.

7.4.1 Gain

The gain may be set to any value up to infinity by a suitable choice of the circulator impedance and the negative resistance of the reflection amplifier. However, amplifiers with gains in excess of 20 dB are likely to be unduly sensitive to changes in circulator impedance with temperature. Also, saturation will occur at a lower input-signal level for a

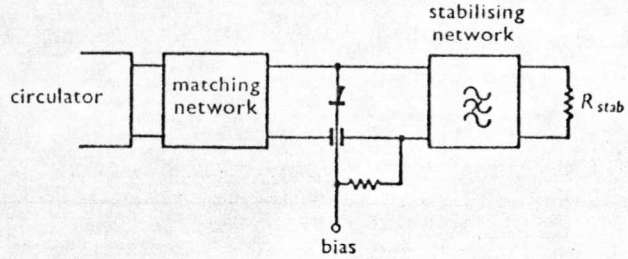


Fig. 35 General form of tunnel-diode amplifier

high-gain amplifier. When a t.d.a. is used as a preamplifier, the gain will be chosen just sufficiently high to render the noise contribution from the second stage negligible. A gain of 15 dB is typical.

7.4.2 Bandwidth

The tunnel diode is an inherently wideband device. Assuming the simplified equivalent circuit of Fig. 31*b*, the gain–bandwidth product for a simple tuning circuit is

$$G^{\frac{1}{2}}B = 1/(\pi RC) \qquad (7.5)$$

and this may be improved by the addition of more reactive elements. The theoretical limit for an infinite number of tuning elements yields the bandwidth

$$B = \frac{4{\cdot}34}{RC(10\log_{10}G)} \qquad (7.6)$$

and this limit is closely approached by six elements.

In practice, the bandwidth is also limited by the diode series resistance and package capacitance C_s. Lead inductance imposes a

further limit on the range of gain (minimum and maximum) realisable for a given bandwidth, and vice versa (Scanlan and Lim, 1965, and Okean, 1966).

Typical figures for an X band germanium tunnel diode are $C_j = 0{\cdot}18$ pF, $R_n = 77\ \Omega$, $C_s = 0{\cdot}25$ pF and resistive cutoff frequency $f_{r0} = 40$ GHz. For an amplifier used at frequencies well below f_{r0}, an effective junction capacitance of $C_j + C_s$ can be assumed, giving the predicted figures shown in Table 4.

The reactance of the circulator may be used to form one or more of the above tuning elements.

Amplifiers with 1 GHz bandwidth at 20 dB gain are available from a number of manufacturers.

Table 4

Gain dB	Bandwidth with simple tuned circuit GHz	Maximum theoretical bandwidth GHz
10	3	13
20	1	6·5

7.4.3 Noise factor

The tunnel diode exhibits shot noise at the junction and also a small amount of thermal noise owing to the bulk series resistance R_s. The latter noise becomes significant only for the higher-frequency amplifiers at X band frequencies and above, where the operating frequency may be in the region of $f_{r0}/3$ or higher.

Nielson (1960) shows the noise factor F to be

$$F = \frac{1 + N_s}{\left(1 - \frac{R_s}{R_n}\right)\left\{1 - \left(\frac{f}{f_{r0}}\right)^2\right\}} \qquad (7{\cdot}7)$$

It is therefore mainly dependent on the noise constant N_s, which is related to the negative characteristic as

$$N_s = \frac{e}{2kT} I_0 R_n \simeq 20 I_0 R_n \quad (\text{at } 290\,^\circ\text{K}) \qquad (7{\cdot}8)$$

where

$$I_0 = \tfrac{1}{2} I_P$$

As the peak current I_P is inversely related to R_n, this is a function of the basic properties of the semiconductor material (Armstrong, 1962), and

$$I_0 \propto \exp A\left\{E_g\left(\frac{\epsilon m}{N}\right)^{\frac{1}{2}}\right\} \qquad (7{\cdot}9)$$

Therefore an appropriate material has a small band gap E_g and a small effective mass m of the lighter carrier, to give a high tunnelling probability. It should also possess a low relative permittivity ϵ, and be capable of being doped to a high level, as N is the weighted average carrier concentration on the two sides of the junction. A high carrier mobility is also desirable, to reduce the series resistance R_s. Table 5 shows the basic material parameters of some of the more promising semiconductors compared with the achieved noise factors and theoretical values, assuming that $f \ll f_{r0}$ and $R_s \ll R_n$.

Table 5

Material	ϵ	E_g eV	m	Electron mobility μ cm²/Vs	N_s	Theoretical noise factor dB	Achieved noise factor dB
Ge	16·0	0·68	0·55	3000	1·28	3·5	4 at 2 GHz 5·5 at 9 GHz
Si	11·8	1·08	1·1	1000	—	—	—
GaAs	12·5	1·40	0·034	5000	1·4	5·5	—
GaSb	14·0	0·70	0·047	4000	0·85	2·5	3·2 at 2 GHz 4·7 at 9 GHz
InSb	17·0	0·18	0·021	100000	0·5*	2	—

* Below about 100 °K

The minimum value of N_s quoted above occurs for a bias voltage somewhat greater than that at the inflection point of the I/V characteristic. Thus, for minimum F, the $R_n C_j$ product is not a minimum, so that bandwidth must be sacrificed if the minimum noise factor is to be obtained. Bandwidth may decrease by 30% as bias is increased from the inflection point to the point of the minimum noise factor.

The intensive study of tunnel-diode design and fabrication techniques, mainly for computer applications, has benefited the

devices for t.d.a.s, and these have virtually achieved the theoretical noise factors.

InSb devices can only be used below 77 °K owing to the low E_g, and little work has been performed on this material (Mukai and Irita, 1967). GaAs, although suitable for use at room temperature, has a poor noise performance. GaAs diodes also suffer some degradation if driven into the thermal-injection region, but this is unlikely to occur under normal amplifier conditions. GaSb offers the lowest noise factor, but with a number of limitations, and therefore Ge probably provides the best overall performance.

7.4.4 Frequency ranges

T.D.A.s are readily obtainable in Britain to a wide variety of specifications between 0·5 and 7 GHz, but there are few suitable diodes for X band frequencies and above, and package inductance and capacitance make design difficult. Above J band, Burrus and Trambarulo (1961) have reported on t.d.a.s operating at frequencies up to 90 GHz with bonded-contact Ge tunnel diodes made in reduced-height waveguide within the amplifier.

7.4.5 Overload characteristics

The peak and valley voltages of a tunnel diode are constant for a given material; so any increase in power handling must be obtained by increasing the peak current. But, as R_n is inversely proportional to peak current, the choice is limited by circuit-design considerations. The majority of t.d.a.s use germanium diodes with a peak current I_p of the order of 2 mA. The input level producing 1 dB gain compression in a single-stage germanium t.d.a. is approximately

$$1 \text{ dB compression point} = -(2G + 20I_p) \text{ dBm} \qquad (7.10)$$

where G is in decibels and I_p is in milliamperes. The 1 dB compression point can be increased by up to 10 dB on selecting an appropriate bias voltage somewhat higher than the I/V inflection point. Thus, a 15 dB-gain amplifier using a 2 mA diode has a 1 dB compression point of −40 dBm when biased for maximum gain and bandwidth at the I/V

inflection point. This can be increased to −33 dBm on increasing the bias voltage. Obtaining the same gain by cascading two low-gain amplifiers gives a further increase in the 1 dB compression point.

One example of obtaining a greater dynamic range without sacrifice in noise factor has been the use of a low-noise Ge tunnel diode followed by a higher-power (and higher-noise) GaAs diode in cascade; this arrangement has been shown by Steinhoff and Stertzer (1964) to result in a 15 dB improvement in dynamic range.

T.D.A.s have also been made using higher-current diodes; a 1 dB compression point of −25 dBm using a 22 mA GaAs diode has been reported. However, package inductance is the main obstacle to the use of high-current tunnel diodes owing to the low values of f_{x0} obtained when R_n is small. An alternative approach by Lee (1967) makes use of two germanium diodes in a push–pull arrangement, which both improves the power-handling capabilities and reduces the intermodulation response of the amplifier.

7.4.6 Distortion

Intermodulation distortion remains very small (signal/distortion > 70 dB) unless the magnitude of the interfering signal is sufficiently large to cause severe gain compression of the amplifier.

A.M.–p.m. conversion can result at large signal levels, but again this effect is small. For example, 0·2 deg/dB a.m.–p.m. conversion occurs at an input level of −20 dBm for a 2 mA germanium diode (Easter, 1965).

7.4.7 Burnout

Permissible c.w. dissipations are of the order of a few tens of milliwatts, but the spike-burnout levels are more difficult to determine, and more extensive tests are required to give quantitative results. American data suggest single-pulse-burnout levels of about 1 erg, but it is difficult to correlate this test with practical conditions and it is significant that the first t.r.-cell–varactor limiters were originally developed in order to protect the t.d.a. Therefore the burnout level of tunnel diodes is probably similar to that of mixer diodes at about 0·1–0·5 erg

for repetitive pulsing. There is also some evidence to suggest that, from this aspect, Ge diodes are superior to either GaAs or GaSb devices, but the main factor is probably the design of the junction element.

7.4.8 Temperature

Ge and GaSb have useful operating temperatures of −55 to 100 °C, whereas GaAs can be used up to 125 °C. This is not a severe restriction in military applications where temperature stabilisation would be employed. However, the peak current of GaSb diodes is more temperature-sensitive and this tends to nullify the slight advantage in noise factor of this material.

7.5 Future trends

The main promise of the t.d.a. is its wide bandwidth, the main difficulty is the low saturation power of the present diodes. Assuming that the package capacitance could be eliminated, Table 4 of Section 7.4.2 could be extended as follows for 20 dB gain:

Table 6

	Bandwidth with single tuned circuit GHz	Maximum theoretical bandwidth GHz
Including C_s	1	6·5
Neglecting C_s	2·3	15·7

So it is clear that improved bandwidth can be obtained by the use of unencapsulated diodes which could be used with other integrated-circuit elements. The decrease of diode inductance thus obtained would also improve circuit stability and enable higher-current tunnel diodes to be used, giving the better power handling required.

Thin-film circuits offer a good initial approach, and Okean (1966*b*) has made a start in this direction on a 4 GHz amplifier. This reflection amplifier has a thin-film tuning inductance, a bypass capacitor, and an

LCR stabilising circuit formed on a small substrate of sapphire. The diode is made by a bonded contact on a small die of germanium, coated with epoxy resin, and subsequently mounted on the substrate.

A considerable improvement in noise factor appears to be unlikely. The noise depends mainly on the noise constant N_s, which is proportional to $I_0 R_n$. The use of higher- or lower-peak-point-current diodes will not alter the noise measure since R_n is inversely proportional to I_0, resulting in a constant value for the product $I_0 R_n$. On employing different semiconductors or doping techniques, N_s may be altered, the effect being essentially to modify the voltage scale of the characteristic. Lower peak-point and valley-point voltages result in a lower N_s, but also a lower dynamic range. The use of GaSb gives a noise factor approximately 0·75 dB lower than that of Ge, but this semiconductor has a less well developed technology and tunnel diodes made from it are reported to be more temperature-sensitive and fragile.

7.6 Summary

The tunnel-diode amplifier is at present used mainly as a simple moderately-low-noise preamplifier for microwave receivers, in cases where the additional complexity of the parametric amplifier is not justified or where it is desirable for the overall noise factor to be dependent on the noise performance of one unit rather than two or more as with a mixer–i.f. chain. The t.d.a. has some power-output limitations, but all low-noise amplifiers have dynamic signal-range restrictions for some system applications, and the t.d.a. has a range of about 60 dB (17 dB gain and 20 MHz bandwidth) and this can be adequate for many applications.

A synthesis procedure is available which enables a stable t.d.a. to be designed with a specified frequency response, taking into account all the circuit elements.

It is possible that, with the advent of better microwave transistors and mixers, the t.d.a. will become obsolete. However, it is evident that some of the potential advantages of the t.d.a. have yet to be realised.

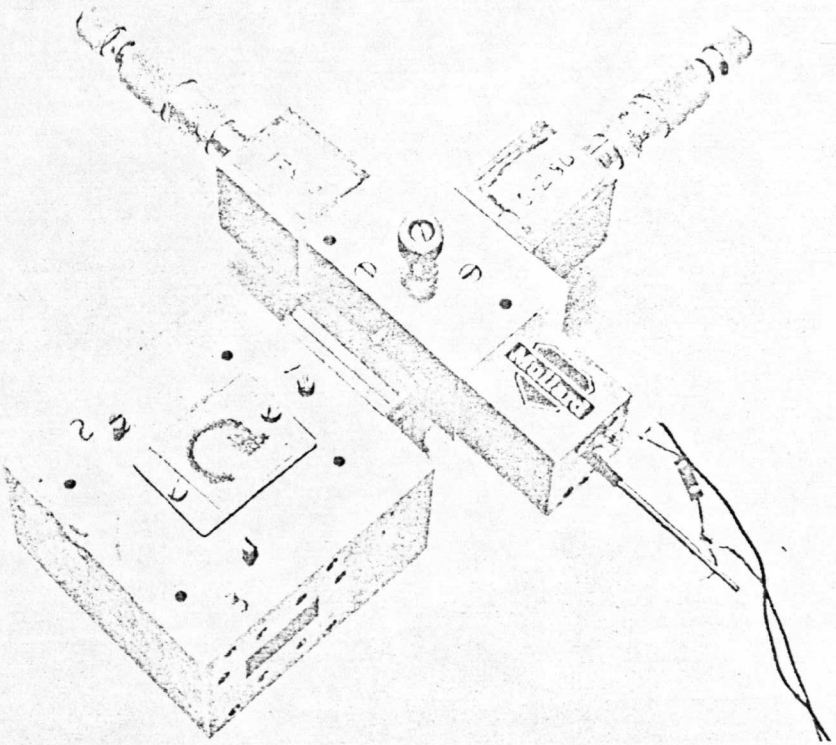
The noise factor of a tunnel diode is governed by the noise constant, cutoff frequency and resistance ratio, and for a germanium device this

puts a theoretical lower limit of about 3·5 dB on the noise factor. The design of these devices is now well understood and noise factors of this order have been achieved at L band and 5·5 dB in X band frequencies.

Operation at a higher current, using a higher-peak-point-current diode, increases the dynamic range but increases the stability problems associated with the higher negative conductance. An alternative approach is to cascade a low-noise Ge t.d.a. with a higher-power GaAs t.d.a.

The stability problems associated with microwave tunnel-diode circuits arise from having a low self-resonant frequency f_{x0}. This frequency is determined principally by the inductance of the diode housing L_s. Further reduction of L_s is very difficult, using diodes in discrete form, but the difficulty would be largely avoided by adopting microcircuit form. Building the amplifier in integrated or thin-film-circuit form may make it possible to reduce stability problems and achieve a higher dynamic range. It is also likely to simplify the design of the associated filter circuits.

Further work on tunnel diodes is needed. At present most alloyed diodes are of etched-mesa construction, somewhat fragile at higher frequencies, and incompatible with integrated circuits. Some progress is being made with planar techniques, the main difficulty being in defining the minute junction area required. However, it now seems feasible to consider an integrated t.d.a. which would offer improved bandwidth and dynamic range. A discussion of microwave amplifiers employing tunnel-diode devices has recently been published by Okean (1967).



18 Complete tunnel-diode amplifier

APPENDIX 14.

A computer programme for the analysis of branched distributed and lumped circuits.

Conference on Electrical Networks, Newcastle upon Tyne, 13th-15th September 1966. I.E.E. Conference Publication No. 23.

A COMPUTER PROGRAMME FOR THE ANALYSIS OF BRANCHED DISTRIBUTED AND LUMPED CIRCUITS

By M.K. McPhun

SUMMARY

A computer programme for predicting the performance of microwave circuits with a small computer, is described. The circuits may be active or passive and may contain combinations of transmission lines, lumped elements, tunnel diodes or other negative resistance devices and series and shunt branches. Loops are not permitted, otherwise circuit complexity is limited only by the storage capacity of the machine. Provision is made for examining circuit sensitivity to variation of any input parameter. Two forms of output are provided: firstly tables v. frequency can be printed for any combination of a number of output options, e.g. input impedance and insertion loss, and secondly, graphs (e.g. v. frequency or Smith charts) can be plotted automatically using a digital plotter.

The programme, written in Algol for an Elliott 503 computer, is in essence a simple compiler. Each type of circuit element is allotted a code number. The data tape contains a list of these code numbers, each followed by the relevant parameter values, the order of the elements on the tape determining the circuit to be analysed. In the programme, each element is represented by its chain matrix, the equivalent chain matrix for the whole network is computed and the output quantities are computed from this equiv alent matrix.

1. INTRODUCTION

This paper is intended for the non-specialist user of a computer for circuit analysis, rather than for the specialist computer programmer.

During the study of microwave tunnel diode amplifiers, a number of programmes were written to evaluate the performance of particular amplifier configurations. This process was found to be most unsatisfactory because any change in the configuration of the amplifier meant changing the programme. It was quickly realised that a programme sufficiently general to handle tunnel diode amplifiers would have many other applications to circuits containing mixtures of distributed and lumped components. It was required to accommodate active and passive circuits containing combinations of transmission lines, lumped elements, tunnel diodes or other negative resistance devices, and series and shunt branches. Loops were not required. It was further desirable that the programme should be capable of easy extension to include extra

---

M.K. McPhun is at the School of Engineering Science, University of Warwick, and was formerly at Mullard Research Laboratories, Redhill, Surrey.

types of circuit elements if they should be needed. Provision for examining circuit sensitivity to variation of any input parameter was necessary.

The form of the programme was influenced by the computer used, an Elliott 503 with a 8,000 word store but no backing store facilities.

Initially output was required in the form of tables vs. frequency for any combination of a number of output options. Later the need for graphical output arose and this was provided using a Calcomp digital plotter. The programme described in sections 2 and 4 is called the Type 1 programme which provides tabulated data only. At the time of writing the graphical options are not fully integrated into a Type 2 programme. The graphical output facilities of the Type 2 programme will be described separately in section 5.

The programme is described from the viewpoint of the user, in section 2. The form of input data and results are given, illustrated by an example. Sections 3 and 4 cover the theory and operation of the programme. The programme is written in Algol; sections 1, 2 and 3 are intended for the general reader without knowledge of Algol, but this will be required for parts of section 4.

2. USE OF THE PROGRAMME

To put a circuit into a suitable form for computation it must be divided into circuit elements. Each type of element is allotted a code number. The data tape contains a list of these code numbers, each followed by the relevant parameter values, and the order of the elements on the tape determines the circuit to be analysed.

2.1 Division of Circuit

The element types available for forming the circuit are listed in Appendix 5; they are of three types. Firstly there are straightforward components such as shunt resistors, transmission lines or tunnel diodes; these components must be given parameter values and are considered as 2 port elements. Secondly there are series and shunt tees which are 3 port elements defining the circuit configuration and possessing no parameters. Thirdly there are one-port branch and final loads which both define the circuit configuration and also possess parameter values. For further illustration we will consider some circuits.

Figure 1 is a simple circuit containing no tees, and is in a suitable form for computation. The generator impedance Zc will be considered in section 2.3. The input data describing the rest of this cireuit is written as table 1.

TABLE 1. Input data describing circuit of figure 1

Number of element	Type of element	Parameter values
1	5	50, 0.1
2	9	20.10^{-12}
3	5	60, 0.2
4	9	30.10^{-12}
5	4, 6	20

Note that the final load must always consist of a shunt R, C or L, so that when element type 4 appears in the table it must be followed immediately by one of the numbers 6, 8 or 10 and then by the appropriate parameter value.

Consider now a circuit requiring the use of tees. Figure 2 shows the circuit of figure 1 with the addition of a lossy inductor and an inductive resistor both shunted across the first capacitor, and with a capacitively loaded transmission line stub added. This circuit is redrawn in figure 3 in a form suitable for writing the input data. The rule to be followed when dealing with tees is to go up each brnach as it occurs, until a branch load is encountered, then return to the preceding tee and continue until the next branch is encountered etc. To each tee there must correspond a branch load so that the number of tees must equal the number of branch loads. The input data of table 2 describing the circuit of figure 3 should make this clear.

TABLE 2. Input data describing the circuit of figure 3

Number of element	Type of element	Parameter values
1	5	50, 0.1
2	2	
3	8	20.10^{-12}
4	1	
5	7	1.0
6	3,10	20.10^{-9}
7	11	10^{-9}
8	3, 6	25
9	5	60, 0.2
10	1	
11	5	20, 0.05
12	3, 8	10^{-11}
13	9	30.10^{-12}
14	4, 6	20

2.2 Variation of a Parameter

The value of any one parameter, including the branch and final load values, can be varied by specifying the initial, incremental and final values in the input data. To identify the varied parameter the trigger 10^{70} is used. For example, suppose in figure 3 the frequency response was required for element no. 9 with values of Zo = 30, 40, 50, 60 and 70 ohms, then the relevant part of table 2 would become as table 3.

TABLE 3.

Number of element	Type of element	Parameter values
"	"	"
"	"	"
8	3, 6	25
9	5	10^{70}, 30, 10, 70, 0.2
10	1	

Note that a small amount must be added to the final value of the varied parameter to ensure that this value is accepted.

2.3 Generator Impedance

For some calculations such as input reflection coefficient the value of generator impedance Zc (figures 1 and 3) is required. - However Zc must always be given a value even if it is not used, as the computer will expect to read a value for it on the data tape (Section 2.5). Normally Zc will be real, but provision is made for inserting a procedure to calculate complex values of Zc that are functions of frequency. To use this facility the value Zc = 10^{-70} is punched in the data tape and the procedure for computing Zc is inserted in the programme when compiling, as described in Section 4.3.

2.4 Output Options

Ten output options are available in any combination. Each option is identified by a number as listed in Appendix 6. The outputs are specified by listing the option numbers, in any order, the list being terminated by the number 0. The definition of these options is given in Section 3.

2.5 Assembly of Complete Data Tape

In addition to the foregoing information it is necessary to state N, the total number of circuits on the data tape, n, the number of elements in the circuit, a title for each set of data, and the frequency range. The data tape must take the form:

N — number of circuits

title
n (number of elements)
Zc
list of output options terminating in 0
table defining circuit, e.g. table 2
start frequency (Gc/s)
incremental frequency (Gc/s)
finish frequency (Gc/s)

} circuit 1

title
n
Zc
'
'
'
'

} circuit 2 etc.

2.5.1 Presentation of Results

It is advantageous to be able to recognise the results without recourse to the data tape; therefore the input data is reproduced with each element identified in words, above the tabulated results. If a parameter is being varied, the current value of this parameter is printed out at the head of each table. Thus it is easy to reconstruct the circuit analysed from the tabulated results.

2.5.2 Example

Figure 4 shows a tunnel diode reflection amplifier. The data tape reproduced in figure 5 is for calculating input impedance, modulus and phase of input reflection coefficient and available reflection gain at frequencies of 0.3 Gc/s, 0.5 Gc/s and 0.7 Gc/s; this to be done for characteristic impedances of the third transmission line of 100Ω, 150Ω and 200Ω. The output obtained is shown in figure 6.

The time taken for this example was 8 seconds for computing and 12 seconds for punching the output tape. In addition the computer takes 40 seconds to digest the owncode version of the translated programme.

3. THEORY

This section is devoted to the circuit analysis method used in the programme. We consider a circuit as built up of a cascade of 2-port circuit elements as in figure 7; circuits containing branches will be considered in Section 3.2. If each circuit element is represented by its "chain" or ABCD matrix $[A_1]$, $[A_2]$ ..... $[A_n]$, then the circuit may be redrawn as in figure 8, where

$$[A] = [A_1]\ [A_2]\ [A_3]\ \ldots\ldots\ [A_{(n-1)}]\ [A_n] \qquad (1)$$

These are all 2 x 2 matrices with complex elements, and Z_c and Z_L may also be complex.

From Z_c, $[A]$, and Z_L all the required quantities can be calculated as follows:

$$\text{Input impedance } Z \text{ in} = \frac{a_{11}\ Z_L + a_{12}}{a_{21}\ Z_L + a_{22}} \text{ ohms} \qquad (2)$$

$$\text{Normalised input impedance} = \frac{Z \text{ in}}{Z_c} \qquad (3)$$

$$\text{Input reflection coefficient } \Gamma = \frac{Zin - Zc}{Zin + Zc} \qquad (4)$$

$$\text{modgamma} = \sqrt{\Gamma .\ \Gamma^*} \qquad (5)$$

$$\text{arggamma} = \tan^{-1} \frac{\text{Im } \Gamma}{\text{Re } \Gamma} \qquad (6)$$

$+180^\circ$ if $\text{Re}(\Gamma) < 0$ and $\text{Im}(\Gamma) > 0$
-180° if $\text{Re}(\Gamma) < 0$ and $\text{Im}(\Gamma) < 0$

The available reflection gain[1] is given by

$$\text{ref Gav} = 10 \log_{10} \left| \frac{Zin - Zc^*}{Zin + Zc} \right|^2 \text{ dB} \qquad (7)$$

The insertion loss is calculated for the general case where neither the generator nor load are matched to each other nor to the network. It is defined as

$$\text{Insertion loss ratio} = \frac{\text{power dissipated in load with network in position}}{\text{power dissipated in load when load is connected directly to the generator}} \quad (8)$$

This is derived in Appendix 1 in a form suitable for computation, and is

$$\text{dB ins} = 10 \log_{10} \left| \frac{(Zc + Zin)\,(a_{21}\,Z_L + a_{22})}{Z_L + Zc} \right|^2 \quad (9)$$

The matrices for the various types of circuit elements are listed in Appendix 2.

3.1 Iterative Solution of Cascaded Circuits

A large storage space would be required to set up all the matrices of equation (1) and then multiply them together. Instead the following method requires space only for [A] and one other matrix, say [B]. Let [A] and [B] also refer to their storage spaces; [A] is set to a unit matrix, the elements of $[A_1]$ are computed and placed in [B], the product of [A]. [B] is formed, and the result placed back in [A]. The elements of $[A_2]$ are then computed and placed in [B] and the process repeated until $[A_n]$ is reached.

3.2 The Treatment of Branches using Series and Shunt Tees

The method used is to replace a branch by an equivalent series or shunt network element in the cascade of figure 7.

Figure 9a shows part of a cascade containing a series branch circuit. This series branch may be represented by an equivalent network with chain matrix $[S] = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$ terminated in a branch load Zb.

The input impedance to the branch Z_s is then given by

$$Z_s = \frac{s_{11}\,Z_b + s_{12}}{s_{21}\,Z_b + s_{22}} \quad (10)$$

and the branch can be replaced by the new circuit element with matrix $[A_k]$ as shown in figure 9b, where

$$[A_k] = \begin{bmatrix} 1 & Z_s \\ 0 & 1 \end{bmatrix}$$

Similarly the shunt branch with input impedance Z_s can be replaced by an equivalent shunt element as figure 9c, with the matrix

$$[A_k] = \begin{bmatrix} 1 & 0 \\ 1/Z_s & 1 \end{bmatrix}$$

In the programme the same sections are used for computing both the input impedance Z_s for the branches and Z_{in} for the main cascade. Also the section which computes the equivalent chain matrix of a cascade is the same whether the cascade occurs in a series or shunt branch, or between the generator and final load. Therefore it is possible to use an iterative method of dealing with branches so that any number of branches and sub-branches can be handled. The way this is done is described in Section 4.

4. DESCRIPTION OF THE PROGRAMME

So far no knowledge of Algol on the part of the reader has been assumed. Algol is a word-orientated language and often any attempt to put an Algol programme into a pictorial form using flow diagrams merely results in extra complication. Therefore Algol language will be used in this section where advantageous. Appendix 7 lists some of the peculiarities of the Elliott version of Algol used for this programme.

The programme is reproduced in Appendix 3. A list of the important integers used in the programme is given in Appendix 4. The flow diagram of figure 10 is intended to show the main processes of the programme but should not be accepted as rigorous. There are four main parts. Firstly, after reading the initial data and output options the table describing the circuit is read. The configuration of the n circuit elements is stored as the order of the circuit elements are read, as described in Section 4.1. The values of the components are temporarily stored in an array V [1 : n, 1 : 10] except for the varied parameter (if any); this is identified and its initial, incremental and final values are stored in U1, U2 and U3.

In the second part the "step parameter" for loop is set up. Once inside this loop the values of the circuit components can be set to the values stored in V [1 : n, 1 : 10], except for the varied parameter which is given its current value. This current value is printed at the head of the results table.

After reading the frequencies required and setting up the "step frequencies" for loop, the third part consists of computing the matrix elements and multiplying the cascades of matrices as described in Section 4.2. Also the circuit branches are sorted out in accordance with the circuit configuration. This is described in Section 4.2.1. At this stage the point is reached where the circuit has been reduced to a generator, an overall equivalent chain matrix, and a load as in figure 8.

The fourth part consists of computing and printing the output data as specified by the output options.

The heart of the programme is the method used for storing the configuration of the circuit, and steering the programme through the circuit element types in the right order. This process is used in the first, second and third parts just

described; the method is ex plained in the next section.

4.2 Storage of Circuit Information

The number of a circuit element is denoted by integer k, where there are n elements in all. As the data tape is read an integer array i |1 : n| is built up, with elements i [k] equal in value to the type number of the relevant circuit elements. Thus for the previous example of figure 4 we have:

Circuit element number k	Circuit type number i [k]
1	i [1] = 5
2	i [2] = 12
3	i [3] = 5
4	i [4] = 6
5	i [5] = 5
6	i [6] = 4

These instructions, together with the instruction goto sss[i[k]] have the label "steer"; sss is a switch list in which the circuit element types are listed and given labels in the same order as their type numbers (c.f. Appendix 5). The particular instructions pertinent to each type of circuit element are followed by the instruction goto steer, so that a closed loop is formed as illustrated in figure 11. This forms block 4 of the programme. Figure 11 also shows that for branch and final loads, an additional circuit element type must be read; this is stored in P |k|. The process is ended when i |k| = 4 (k then equals n) at the final load and P |k| for the final load has been read.

At this stage the circuit component values are read into an array V |0 : n, 0 : 10| by the instruction read V |k, Q|. The upper limit of Q depends on the circuit type, e.g. for the shunt tunnel diode Q goes from 1 to 5 whereas for the shunt resistor Q = 1 only. If V |k, Q| > 9.10^{69} then this signifies the varied parameter and its initial, incremental and final values are read into U1, U2 and U3. At the same time the current values of k and Q are placed in k1 and B respectively so that V |k1, B| will identify which variable must take the values U1, U2, U3 in the subsequent "step parameter" for loop.

Also at this stage the tees and branch loads are counted, but we will postpone further discussion of this until they are treated in detail in Section 4.2.1.

4.2 Computation of the Equivalent Overall Matrix

The principle of the iterative calculation was discussed in Section 3.1. This process takes place in the "step frequency" for loop. Complex algebra is used with, for instance, the real part stored in 'a' and the imaginary part in jb; the 'j' has no significance except to enable the imaginary part to be recognised.

Procedure mult ( ) and div ( ) enable the complex numbers to be multiplied and divided whilst procedure chain ( ) enables the product [A].[B] = [C] to be formed where [A], [B] and [C] are 2 x 2 complex matrices. An important feature of all these

procedures is that the result can be put back into one of the original locations, i.e. [A]: = [A].[B]. The complex algebra procedures of van de Kiet[2] would have given more compact expressions, but it was considered that the method used here is easier to follow.

A very similar process to that described in Section 4.1 is used for guiding the computer through the network; here the guiding is done by the instruction labelled "directory" which contains the instructions:

```
chain (A, jB, a, jb,  A, jB);
k:  = k + 1;
goto  s s [i[k]];.
```

The switch list s s here contains the calculations for the elements of all the matrix types. After computing the matrix elements the instruction "goto directory" repeats the process until the final load is reached. Special action is, of course, required for the tees which we will now consider.

4.2.1. Computation of Branches

When the circuit configuration is read, as described in Section 4.1, the total number of tees in the circuit is counted using integer h. This number is then stored in integer h1. h1 is then used later, before the calculations are commenced in the "step frequency" for loop, to reserve an adequate amount of storage space in the array A1,jB1[1:2*h,1:2]

Referring to figure 12, the computations when tees are present are as follows. h is again used to count each tee as it is encountered, but now h: = h-1 at a branch load, so that h effectively remembers which branch the programme is on at any moment. Also, a record is made of whether the branch starts with a series or a shunt tee by setting the integer array m [h] : = 1 for a shunt tee, m [h] : = 2 for a series tee.

The chain matrix calculated for the main cascade of figure 12 is in array A, jB, and remains in this location all the time that h = 0. When the first tee is encountered, the array A, jB is transferred for storage in array A1, jB1; array A, jB is reset to give a unit matrix and a new chain multiplication is commenced up the branch. If another tee is encountered up this branch, then array A, jB is again stored in A1, jB1, a new set of locations being selected using the integer j = 2h.

When a branch load is reached the input admittance of the branch is computed using procedure Y in ( ). Then for the shunt tee (m [h] = 1) the elements of the equivalent shunt matrix are formed using procedure shuntset ( ), while for a series tee (m [h] = 2) the elements of the equivalent series matrix are formed using procedure series set ( ). The principle of these calculations were explained in Section 3.2.

The equijvalent array for theprevious branch, stored in A1, jB1, can now be recovered and placed back in A, jB and the calculations for the previous branch be proceeded with.

4.3 Additional Facilities

As described in Section 2.3 it is possible to insert a procedure for computing a generator impedance that is a function of frequency. When a value of $Z_c < 10^{-60}$ is inserted in the data, the programme prints "procedure" instead of the value in the headings to the results. The procedure zedsee (x, jy) must be inserted in the programme at the place indicated immediately before computing the output options; it must be followed by the procedure call zedsee (re Z_c, imZ_c).

4.3.1 Extension to include New Types of Elements

It is very simple to include new types of elements. Provision has been made for a series, and a shunt element to be added as types 20 and 21. To do this three things must be done; instructions for reading and printing the circuit values must be inserted at labels L20 and L21 in block 3, the values of the circuit components must be settat labels L20 and L21 in the "step parameter" for loop, and the appropriate calculations must be inserted at labels s20 and s21 in the "step frequency" for loop.

The elements type 18 and 19 are for reading tables of admittance and impedance versus frequency. This facility has not yet been included.

5. GRAPHICAL OUTPUT

Options 11 to 14 have been added to give graphical presentation of output. Examples of cartesian plots v. frequency and Smith chart plots will be presented at the conference.

6. CONCLUSIONS

Though the programme was written for the analysis of tunnel diode reflection amplifiers, it was made sufficiently general to be applied to a wide variety of circuits. The circuits may consist of combinations of transmission lines, lumped elements, tunnel diodes or other negative resistance devices, and series and shunt branches, but loops are not permitted. Circuit complexity is limited only by the storage capacity of the machine, and this limit has not been met in practice so far. The provision made for examining circuit sensitivity to any parameter has been found most useful.

The method of setting out input data, described in Section 2, has been found to be quite convenient. This enables the non-specialist to use the programme with ease.

Some aspects of the programme could be more satisfactory, the locations occupied by the large array V [0 : n, 0 : 10] cannot be used again in the later stages of the programme, and no way has been found of making this possible. It would also be desirable to use the main body of the programme for computing Z_c when it is a function of frequency.

At present the time taken for punching out results at 100

characters/sec. is usually roughly equal to the computing time, so that a buffer store for the punch would halve the running time.

Great care was taken in labelling the results (c.f. example of figure 6). We believe that this has been well worth while. This has saved much time as any doubt about the input conditions pertaining to computed tables of results renders them quite useless.

The programme has proved to be a valuable facility in the laboratory.

7. ACKNOWLEDGMENT

Many helpful discussions on the use of Algol were held with Mr. K. Tweedale.

8. REFERENCES

1. Kurokawa, K., "Power Waves and the Scattering Matrix", I.E.E.E. Trans, 1965, MTT-13, p. 194.
2. van de Riet, "Complex Arithmetic", comm. A.C.M. 7, July, 1963, Algorithm No. 186.

APPENDIX 1

Calculation of Insertion Loss

With reference to Fig. 13, the chain matrix is defined as

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$$

giving $V_1 = a_{11} V_2 + a_{12} I_2$ (A 1)

$I_1 = a_{21} V_2 + a_{22} I_2$ (A 2)

also $V_2 = Z_L \cdot I_2$

substituting in equation (A 2), we have

$I_1 = I_2 (a_{21} Z_L + a_{22})$ (A 3)

For the definition of equ 8 we wish to calculate the power dissipated in the load, with and without the network inserted.

In general, the power dissipated in load is

$$P_L = |I_2|^2 \cdot \mathrm{Re}\ (Z_L) \quad \text{(A 4)}$$

With network,

$$P_L = \left| \frac{I_1}{(a_{21} Z_L + a_{22})} \right|^2 \mathrm{Re}\ (Z_L) \quad \text{(A 5)}$$

where $I_1 = \dfrac{V_g}{Zc + Zin}$

$$\therefore \quad P_L = \left| \frac{V_g}{(Zc + Zin)(a_{21} Z_L + a_{22})} \right|^2 \mathrm{Re}\ (Z_L) \quad \text{(A 6)}$$

Without the network, power P_o in load is

$$P_o = \left| \frac{V_g}{Z_L + Zc} \right|^2 \mathrm{Re}\ (Z_L) \quad \text{(A 7)}$$

Insertion loss ratio

$$\frac{P_o}{P_L} = \frac{\left| \dfrac{V_g}{Z_L + Zc} \right|^2}{\left| \dfrac{V_g}{(Zc + Zin)(a_{21} Z_L + a_{22})} \right|^2}$$

$$\text{Insertion loss} = 10 \log_{10} \left| \frac{(Zc + Zin)(a_{21} Z_L + a_{22})}{Z_L + Z_c} \right|^2 \quad \text{(A 8)}$$

APPENDIX 2

Chain Matrices for the Circuit Elements

Element Type No.	Label	Diagram	Formulae
5	line	(a)	$\beta = \frac{\omega}{c}$
6	shunt R	(b)	$Y = {}^{I}/R$
7	series R	(c)	$Z = R$
8	shunt C	(b)	$Y = j\ \omega\ Cs$
9	series C	(c)	$Z = {}^{1}/j\ \omega\ Cs$
10	shunt L	(b)	$Y = {}^{1}/j\ \omega\ Ls$
11	series L	(c)	$Z = j\ \omega\ Ls$
12	shunt T.D.	(b) also Fig.15	$Y = j\ \omega\ C1 + {}^{1}/Z'$ where $Z' = r - \frac{G}{G^2 + \omega^2 c^2} + j\left(\omega L - \frac{\omega C}{G^2 + \omega^2 c^2}\right)$ and G is a positive number
13	series T.D.	(c) also Fig.15	$Z = {}^{1}/$(Y for type 12)
14	shunt sc stub	(b)	$Y = {}^{1}/jZo \tan(\omega\ \ell/c)$
15	series sc stub	(c)	$Z = j\ Zo \tan(\omega\ \ell/c)$
16	shunt oc stub	(b)	$Y = j\ Zo \tan(\omega\ \ell/c)$
17	series oc stub	(c)	$Z = {}^{1}/j\ Zo \tan(\omega\ \ell/c)$

APPENDIX 3

The Programme

```
M.K. McPHUN   MICROWAVE CIRCUIT ANALYSIS,TYPE 1 ;

begin comment block 1 ;
      integer N,case ;
      read N ; for case:= 1 step 1 until N do
      begin comment block 2 ;
      integer n,t ; integer array opt[0:10] ; real Zc ;
            begin comment read title ; integer array A[1:100] ; integer m ;
            m:= 1 ; instring (A,m) ; m:=1 ; outstring (A,m) ;
            end of title ;
      read n,Zc ; print ££1?n=?,digits(2),sameline,n,££s6?Zc=? ;
      if Zc>10-60 then print sameline, freepoint(4), Zc, £ohms?
                  else print £procedure?;  print ££1?Output options?;

            begin comment read and set output options ;
                  integer p,t ; t:= 1 ;
                  for p:=1 step 1 until 10 do opt[p]:= 0 ;
                        for p:=1 step 1 until 10 do
                        begin if t≠0 then begin read t ;
                                    print sameline,digits(2),t ;
                                          opt[t]:= 1 ;
                                          end
                        end
            end of output options ;

      begin comment block 3 ; integer k,k1,B,h,h1,j,B1 ;
      switch S:= read frequencies, step parameter ;
      integer array i[1:n],P[1:n] ;
      array Zo,l,Rs,Cs,Ls,R,C,r,L,C1[1:n],
            V[0:n,0:10] ;
      real U1,U2,U3,f1,f2,f3,f;
      B:=0 ; k1:= 0 ; h:= 0 ;B1:=0 ;
            begin comment block 4, read in input data for case and print it ;
            integer Q ;
            switch s:= steer,M1,M2,M3,M4,M5,M6 ;
            switch sss:= L1 shunt tee, L2 series tee, L3 branch load, L4 final load,
                        L5 line,L6 shunt R,L7 series R,L8 shunt C,L9 series C,
                        L10 shunt L, L11 series L,L12 shunt TD,L13 series TD,
                        L14 shunt sc stub, L15series sc stub,L16 shunt oc stub,
                        L17 series sc stub, L18 shunt table,L19 series table,L20
                        shunt element,L21 series element ;
            procedure read ;
                  begin read V[k,Q] ; if V[k,Q]>9₁₀69 then
                        begin k1:= k ; B:=B1:= Q ; read U1,U2,U3;
                              V[k,Q]:= U1 ;
                        end V[k1,B]will have to take the values U1,U2,U3
                              in an overall cycle
                  end of read ;
            procedure print 1 ;
                  begin print digits(2),k,££s4??,sameline,i[k],££s2??
                  end of print 1 ;
            procedure print 2 ;
                  begin if B1=Q then print sameline,scaled(4),U2,U3 ;
                        Q:= Q+1 ; B1:=0 ;
                  end of print 2 ;
```

```
            print££12?? ;
steer:            read k,i[k] ;
            goto sss[i[k]] ;
L1 shunt tee: print 1 ; h:= h+1 ;
            print £shunt tee? ;
            goto steer ;

L2 series tee: print 1 ; h:= h+1 ;
            print £series tee? ;
            goto steer ;

L3 branch load:
            print 1 ; print £branch load? ;
M1:         read P[k] ; print sameline,digits(2),££s2??,P[k] ,££s2?? ;
            goto sss [P[k]] ;
            comment this leaves load element value in V[k,Q] while P[k] identifies
                  the type of element ;

L4 final load:
            print 1 ; print £final load? ;
            goto M1 ;

L5 line :   print 1 ; print £line? ;
M6:         Q:= 1 ; read ; print££1t5?Zo=?,freepoint(3),sameline,V[k,Q] ;
            print 2 ; print £ohms? ;
                  read ; print££1t5s?1=?,freepoint (4),sameline,V[k,Q] ;
            print 2 ; print £m? ;
            goto steer ;

L6 shunt R:       if i[k]≠3 and i[k]≠4 then print 1 ;
            print £shunt R? ;
M2:         Q:= 1 ; read ; print££1t5?Rs=?,freepoint (4),sameline,V[k,Q] ;
            print 2 ; print £ohms? ;
            if i[k]=4 then goto read frequencies ;
            goto steer ;

L7 series R:
            print 1 ; print £series R? ;
            goto M2 ;

L8 shunt C:       if i[k]≠3 and i[k]≠4 then print 1 ;
            print £shunt C? ;
M3:         Q:= 1 ; read ; print££1t5?Cs=?, scaled (4),sameline,V[k,Q] ;
            print 2 ; print £F? ;
            if i[k]=4 then goto read frequencies ;
            goto steer ;

L9 series C:      print 1 ; print £series C? ;
            goto M3 ;

L10 shunt L:      if i[k]≠3 and i[k]≠4 then print 1 ;
            print £shunt L? ;
M4:         Q:= 1 ; read ; print££1t5?Ls=?,scaled (4),sameline,V[k,Q] ;
                  print 2 ; print £H? ;
            if i[k]=4 then goto read frequencies ;
            goto steer ;
```

```
L11 series L:print 1 ; print £series L? ;
            goto M4 ;

L12 shunt TD:
            if i[k]≠3 and i[k]≠4 then print 1 ;
            print £shunt T.D.? ;
M5:         Q:= 1 ; read ; print££lt5s?R=?,freepoint(4),sameline,V[k,Q] ;
            print 2 ; print £ohms? ;
            read ; print££lt5s?C=?,scaled(4),sameline,V[k,Q] ;
            print 2 ; print £F? ;
            read ; print££lt5s?r=?,freepoint(4),sameline,V[k,Q] ;
            print 2 ; print £ohms? ;
            read ; print££lt5s?L=?,scaled(4),sameline,V[k,Q] ;
            print 2 ; print £H? ;
            read ; print££lt5?Cs=?,scaled(4),sameline,V[k,Q] ;
            print 2 ; print £F? ;
            if i[k]=4 then goto read frequencies ;
            goto steer ;

L13series TD:
            print 1 ; print £series T.D.? ;
            goto M5 ;

L14 shunt sc stub:
            print 1 ; print £shunt sc stub? ;
            goto M6 ;

L15 series sc stub:
            print 1 ; print £series s/c stub? ;
            goto M6 ;

L16 shunt oc stub :
            print 1 ; print £shunt oc stub? ;
            goto M6 ;

L17 series oc stub :
            print 1 ; print £series o/c stub? ;
            goto M6 ;

L18 shunt table :
            print 1 ; print £shunt table? ;
            goto steer ;

L19 series table :
            print 1 ; print £series table? ;
            goto steer ;

L20 shunt element :
            print 1 ; print £shunt element? ;
            goto steer ;

L21 series element:
            print 1 ; print £series element?  ;
            goto steer ;


            end of block 4 ;
```

```
read frequencies:
     read f1,f2,f3 ; print££13?? ;  h1:=h ;
     if B=0 then     begin U1:=0 ;  U2:=2 ;  U3:=1 ;
                       end;
     comment this is the case if no parameter is to be varied. The following
          for list sets the appropriate parameter  ;
step parameter:

     for V[k1,B]:= U1 step U2 until U3 do
          begin
               comment in addition to setting the variables,
                    this block prints the value of the varied parameter
                    (if any) at the head of the results table ;
                         switch ssss:= END;
          switch s:= L1 shunt tee,L2 series tee, L3 branch load, L4 final load,
                    L5 line, L6 shunt R, L7 series R, L8 shunt C, L9 series C,
                    L10 shunt L, L11 series L, L12 shunt TD, L13 series TD,
                    L14 shunt sc stub, L15 series sc stub, L16 shunt oc stub,
                    L17 series oc stub, L18 shunt table, L19 series table,
                    L20 shunt element, L21 series element ;


               for k:= 1 step 1 until n do
                    begin digits (2) ; sameline ;
                    goto s[i[k]] ;
L1 shunt tee:        goto END;
L2series tee:       goto END;
L3 branch load:     goto s[P[k]] ;
                    comment this will allocate appropriate type of variable
                         to V[k,0] ;

L4 final load:      goto s[P[k]] ;

L5 line:            Zo[k]:= V[k,1] ; l[k]:= V[k,2] ;
                    if k=k1 then
                         begin if B=1 then print £Zc?,k,£=?,freepoint(4),Zo[k]
                              ,£ohms?
                         else if B=2 then print £l?,k,£=?,
                              scaled(4),l[k],£m? ;
                         end;
                         goto END;

L6 shunt R:         Rs[k]:=V[k,1] ;
               if k=k1 then print £Rs?,k,£=?,freepoint(4),Rs[k],£ohms? ;
                         goto END;

L7 series R :    goto L6 shunt R ;

L8 shunt C:         Cs[k]:=V[k,1] ;
               if k=k1 then print £Cs?,k,£=?,scaled(4),Cs[k],£F? ;
                         goto END;

L9 series C:     goto L8 shunt C ;

L10 shunt L:     Ls[k]:=V[k,1] ;
          if k=k1 then print £Ls?,k,£=?,scaled(4),Ls[k],£H? ;
               goto END;
```

```
L11 series L:                goto L10 shunt L ;

L12 shunt TD:          R[k]:= V[k,1] ; C[k]:= V[k,2] ; r[k]:= V[k,3] ;

          L[k]:= V[k,4] ; C1[k]:= V[k,5] ;
          if k=k1 then
                begin switch s:= L1R,L2C,L3r,L4L,L5C1,E ;
                goto s[B] ;
                L1R: print £R?,k,£=?,freepoint(4),R[k],£ohms? ;
                     goto E ;
                L2C: print £C?,k,£=?,scaled(4),C[k],£F? ;
                     goto E ;
                L3r: print £r?,k,£=?,freepoint(4),r[k],£ohms? ;
                     goto E ;
                L4L: print £L?,k,£=?,scaled(4),L[k],£H? ;
                     goto E ;
                L5C1: print £C1 ,k,£=?,scaled(4),C1[k],£F? ;
                E: end of shunt T.D. ;
                goto END;

L13 series TD:         goto L12 shunt TD ;

L14 shunt sc stub:                  goto L5 line ;

L15 series sc stub:    goto L5 line ;

L16 shunt oc stub:      goto L5 line ;

L17 series oc stub:    goto L5 line ;

L18 shunt table:       goto END;
L19 series table:      goto END;

L20 shunt element:     goto END;

L21 series element:    goto END;

END:             end of setting variables ;
          print ££12?? ;

          comment Now print titles;

          print £Freq£s2?? ;
          if opt[1]=1 then print££s2?reZin£s3?imZin£s?? ;
          if opt[2]=1 then print££s2?renZin£s2?imnZin? ;
          if opt[3]=1 then print££s2?reYin£s3?imYin£s?? ;
          if opt[4]=1 then print££s2?renYin£s2?imnYin? ;
          if opt[5]=1 then print££s2?reZc£s4?imZc£s2?? ;
          if opt[6]=1 then print££s2?modgam? ;
          if opt[7]=1 then print££s2?arggam? ;
          if opt[8]=1 then print££s2?refGav? ;
          if opt[9]=1 then print££s2?reYD£s4?imYD£s2?? ;
          if opt[10]=1 then print££s2?dBins? ;

          print ££1?(Gc/s)? ;
```

```
            if opt[1]=1 then print££s2?(ohms)£s2?(ohms)? ;
            if opt[2]=1 then print££s16?? ;
            if opt[3]=1 then print££s2?(mmho)£s2?(mmho)? ;
            if opt[4]=1 then print££s16?? ;
            if opt[5]=1 then print££s2 (ohms)£s2?(ohms)? ;
            if opt[6]=1 then print££s8?? ;
            if opt[7]=1 then print££s2?(deg)£s?? ;
            if opt[8]=1 then print££s2?(dB)£s2?? ;
            if opt[9]=1 then print££s2?(mmho)£s2 (mmho)? ;
            if opt[10]=1 then print££s2?(dB)£s2?? ;
prefix (££1??) ;



            for f:= f1 step f2 until f3 do
                  begin comment step frequency ;
                  integer p,q ; integer array m[0:h1]  ;
                  switch s:= directory ;
                  array A,jB,a,jb[1:2,1:2],A1,jB1[1:2*h1,1:2] ;
                  real RL,jXL,reZin,imZin,renZin,imnZin,reYin,imYin,
                        renYin,imnYin,reZc,imZc,refGav,dBins,K,
                        modgam,arggam,reYB,imYB,reYD,imYD,w ;
                  switch ss:= s1 shunt tee, s2 series tee, s3 branch load,
                        s4 final load, s5 line, s6 shunt R, s7 series R,
                        s8 shunt C, s9 series C, s10 shunt L, s11 series L,
                        s12 shunt TD, s13 series TD, s14 shunt sc
                        stub, s15 series sc stub, s16 shunt oc stub,
                        s17 series oc stub, s18 shunt table, s19 series table,
                        s20 shunt element, s21 series element ;
      procedure mult(a,jb, c,jd, e,jf);   real a,jb, c,jd, e,jf;
            comment this procedure multiplies the complex number a+jb by c+jd
                  and puts the result into e+jf.  When calling this procedure the
                  result may be put back into a+jb or c+jd if desired;
            begin real x;     x:=a*c - jb*jd;  jf:=a*jd + jb*c;  e:=x;
            end of mult;
procedure chain(a,jb, c,jd, e,jf);   array a,jb, c,jd, e,jf;
comment a+jb and c+jd are 2 by 2 complex matrices.  This procedure multiplies them
      together and puts the result into e+jf.  For use in multiplying chains of
      matrices the result may be put into a+jb or c+jdif desired when the procedure
      is called;
      begin integer i,j,s;real S,jS;  array x,jy[1:2,1:2];
            for i:=1,2 do    for j:=1,2 do
                  begin real P,jP;   S:=0;   jS:=0;
                        for s:=1,2 do
                              begin mult(a[i,s],jb[i,s], c[s,j],jd[s,j], P,jP);
                                    S:= S+P ; jS:= jS+jP
                              end;
                              x[i,j]:=S;   jy[i,j]:=jS;
                  end of partial sum;
            comment it remains to transfer the elements x+jy to e+jf;

            for i:=1,2 do  for j:=1,2 do
                  begin e[i,j]:=x[i,j];    jf[i,j]:=jy[i,j]
                  end
      end of chain;
```

```
procedure div(a,jb, c,jd, e,jf);  real a,jb, c,jd, e,jf;
comment this procedure divides the complex number a+jb by c+jd and puts the
        result into e+jf.When calling this procedure the result may be put back
        into a+jb or c+jd if desired;
      begin real x,y;  x:=c*c+jd*jd;  y:=(a*c+jb*jd)/x;  jf:=(jb*c-a*jd)/x;
                e:=y
      end of div;

      procedure shuntset (G,jB) ; real G, jB ;
      comment this sets the elements of matrix a,jb for a shunt circuit ;
            begin jb[1,1]:= jb[1,2]:= jb[2,2]:=  a[1,2]:= 0 ;
                  a[1,1]:= a[2,2]:= 1 ;
                  a[2,1]:= G ; jb[2,1]:= jB ;
            end of shuntset ;


      procedure seriesset (X,jY) ;  real X,jY ;
            comment this sets the elements of matrix a,jb for a series circuit;
            begin jb[1,1]:= jb[2,1]:= jb[2,2]:= a[2,1]:= 0 ;
                  a[1,1]:= a[2,2]:= 1 ;
                  a[1,2]:= X ; jb[1,2]:= jY ;
            end of seriesset ;
      procedure Yin(reY,imY) ; real reY,imY ;
            comment this computes the input admittance for the matrix A,jB
                  terminated in various loads, and the load impedances for
                  those loads ;
            begin real a,jb,c,jd ; switch s:= E ;
                  if P[k]=6 then begin RL:=Rs[k] ;
                              jXL:= 0 ;
                              end of shunt R load
                  else if P[k]= 8 then
                        begin if Cs[k]= 0 or w= 0 then
                              begin div (A[2,1],jB[2,1],
                                    A[1,1],jB[1,1],reY,imY) ;
                                     goto E ;
                                     end ;
                              RL:= 0 ; jXL:=- 1/(w*Cs[k]) ;
                              end of shunt C load
                  else if P[k]= 10 then
                        begin RL:= 0 ;
                              jXL:= w*Ls[k] ;
                  end of shunt L load
            else if P[k]= 18 then begin
                              end of shunt table load
            else if P[k]= 20 then begin
                              end of shunt element
                                 load ;
            mult (A[2,1],jB[2,1],RL,jXL,a,jb) ;
            mult(A[1,1],jB[1,1],RL,jXL,c,jd) ;
            div((a+A[2,2]),(jb+jB[2,2]),(c+A[1,2]),(jd+jB[1,2]),reY,imY) ;
      E:    end of procedure Yin ;
      w:= 2*3141592600*f ; K:= 1/299850000 ;
      comment set matrix A,jB to a unit matrix, set elements
            of the other matrices in order, and multiply the chain ;
      A[1,1]:= A[2,2]:= 1 ;
      A[1,2]:= A[2,1]:= jB[1,1]:= jB[2,2]:= jB[1,2]:= jB[2,1]:= 0 ;
      h:= 0 ; k:= 1 ;
      goto ss[i[k]] ;
```

```
          comment ss[i[k]] sets the matrix elements, and, except
                for tees and branch loads, returns to directory ;


directory:      chain(A,jB,a,jb,A,jB) ;
          k:= k+1 ; goto ss[i[k]] ;
s1 shunt tee:
          if m[h]≠2 then m[h]:= 1 ; j:= 2*h ;
          for p:= 1,2 do
                for q:= 1,2 do
                begin comment store the current value of matrix A,jB in A1,jB1 ;
                      A1[p+j,q]:= A[p,q] ;
                      jB1[p+j,q]:= jB[p,q] ; jB[p,q]:= 0 ;
                end now set matrix A,jB to unity ;
          A[1,2]:= A[2,1]:= 0 ; A[1,1]:= A[2,2]:= 1 ;
          h:= h+1 ; k:= k+1 ;
          goto ss[i[k]] ;
          comment this sends the programme up a network branch,
                storing the value that matrix A,jB had when
                entering the tee, in A1,jB1 ;
s2 series tee:
          m[h]:= 2 ; goto s1 shunt tee ;
s3 branch load:
          h:= h-1 ; j:= 2*h; Yin(reYB,imYB) ;
          if m[h]=1 then  shuntset (reYB,imYB)
          else if m[h]=2 then begin real reXB, imXB ;
                                    div (1,0,reYB,imYB,reXB,imXB) ;
                                    seriesset (reXB,imXB) ;
                                    end ;
          comment now replace matrix elements stored in A1,jB1 ;
          for p:= 1,2 do
                for q:= 1,2 do
                      begin A[p,q]:= A1[p+j,q] ;
                            jB[p,q]:= jB1[p+j,q] ;
                      end ;
          goto directory ;
s5 line:        begin real x,y ;
                jb[1,1]:= jb[2,2]:= a[1,2]:= a[2,1]:= 0 ;
                x:= K*l[k]*w ;
                a[1,1]:= a[2,2]:= cos(x) ;  y:= sin(x) ;
                jb[1,2]:= Zo[k]*y ; jb[2,1]:= y/Zo[k] ;
          end of line ;
          goto directory ;
s6 shunt R;
          shuntset (1/Rs[k],0) ;
          goto directory ;
s7 series R:    seriesset (Rs[k],0) ; goto directory ;
s8 shunt C:     shuntset (0,w*Cs[k]) ; goto directory ;
s9 series C:    seriesset (0,-1/(w*Cs[k])) ; goto directory ;
s10 shunt L:    shuntset (0,-1/(w*Ls[k])) ; goto directory ;
s11 series L: seriesset (0,w*Ls[k]) ; goto directory ;
s12 shunt TD:
          begin real x,jy ;
                div (1,0,(-1/R[k]),w*C[k],x,jy) ;
                div (1,0,(r[k]+x),(w*L[k]+jy),reYD,imYD) ;
          imYD:=imYD+w*C1[k];shuntset(reYD,imYD);
          end of shunt T.D. ;  goto directory ;
```

```
s13 series TD:
                begin real x,jy ;
                      div (1,0,(-1/R[k]),w*C[k],x,jy) ;
                      div (1,0,(r[k]+x),(w*L[k]+jy),reYD,imYD) ;
                      imYD:=imYD + w*C1[k] ;
                div (1,0,reYD,imYD,x,jy) ;
                      seriesset (x,jy) ;
                end of series T.D. ; goto directory ;
s14 shunt sc stub:
          shuntset (0,(-1/(Zo[k]*tan(K*l[k]*w)))) ; goto directory ;
s15 series sc stub:
          seriesset (0,(Zo[k]*tan(K*l[k]*w))) ; goto directory ;
s16 shunt oc stub:
          shuntset (0,(tan(K*l[k]*w))/Zo[k]) ; goto directory ;
s17 series oc stub:
          seriesset(0,(-Zo[k]/tan(K*l[k]*w))) ; goto directory ;
s18 shunt table:
          goto directory ;
s19 series table:
          goto directory ;
s20 shunt element:
          goto directory ;
s21 series element:
          goto directory ;
s4 final load: Yin(reYin,imYin) ;
          comment this procedure, besides computing Yin for the
                whole network, will have computed the
                appropriate values of RL and jXL for the load;
                     begin own  integer i;
                           if f=f1 then i:=0;
                           i:=i+1;
                           if i=5 then
                                 begin print££1??;
                                 i:=0;
                                 end
                     end;



          freepoint(4) ;
          print f ;
          prefix(££s2??) ;


          comment now compute Zin, Zc if necessary, and nZin, then
                the other output quantities are computed only if
                called for in the output options ;
          div(1,0,reYin,imYin,reZin,imZin) ;
          if Zc<10-60 then
          begin  comment procedure zedsee(x,jy) to
                      be inserted here when appropriate. It
                      must be followed by the procedure call zedsee(reZc,imZc) ;
                div(reZin, imZin, reZc, imZc, renZin, imnZin) ;
          end  .
```

```
        else begin renZin:= reZin/Zc ; imnZin:= imZin/Zc ;
             end ;
        if opt[1]=1 then print reZin,imZin ;
        if opt[2]=1 then print renZin, imnZin ;
        if opt[3]=1 then print reYin*1000, imYin*1000 ;
        if opt[4]=1 then
             begin div (1,0,renZin,imnZin,renYin,imnYin) ;
                   print renYin, imnYin ;
             end ;
        if opt[5]=1 then begin if Zc<10-6 then print reZc,imZc
                         else print Zc,ff8?? ;
             end ;
        if opt[6]=1 then
             begin real regamma,imgamma ;
                   div ((renZin-1),imnZin,(renZin+1),imnZin,
                        regamma,imgamma) ;
                   modgam:=sqrt( regamma*regamma+imgamma*imgamma) ;
                   print modgam ;
                   arggam:= (arctan(imgamma/regamma))*180/3.1415927 ;
                   if regamma < 0 then
                        begin if imgamma > 0
                              then arggam:= arggam+180
                              else arggam:= arggam -180 ;
                        end
                   end ;
        if opt[7]=1 then print arggam ;
        if opt[8]=1 then
             begin real a,jb ;
             if Zc>10-6 then div ((renZin-1), imnZin,
                                  (renZin+1),imnZin,a,jb)
                   else div ((reZin-reZc),(imZin+imZc),
                             (reZin+reZc),(imZin+imZc),a,jb) ;
             refGav:= 4.343*ln(a*a+jb*jb) ;
             print refGav ;
             end  ;
        if opt[9]=1 then print reYD*1000,imYD*1000 ;
        if opt[10]=1 then
             begin  real a,jb;
                    mult (A[2,1],jB[2,1],RL,jXL,a,jb) ;
        if Zc<10-60 then
             begin mult ((reZc+reZin),(imZc+imZin),
                         (a+A[2,2]),(jb+jB[2,2]),a,jb) ;
             div(a,jb,(RL+reZc),(jXL+imZc),a,jb) ;
             end
        else begin  mult ((Zc+reZin),imZin,(a+A[2,2]),
                                  (jb+jB[2,2]),a,jb) ;
             div(a,jb,(RL+Zc),jXL,a,jb) ;
             end ;
        dBins:=4.343*ln(a*a+jb*jb) ;
        print dBins ;
             endof opt[10] ;
             prefix(ff1??) ;
        end of step frequency ;
   print ff14?? ;
   end of step parameter
   end of block 3 ;
   print ff110?? ;
   end of block 2
end of program ;
```

APPENDIX 4

Important Integers Used in the Programme

N	Total number of circuits to be analysed
case	number of particular circuit (= 1, 2, ... N)
n	total number of elements in circuit
k	position of particular element (= 1, 2, ... n)
k1	remembers value of k for the element containing the varied parameter
i [k]	type number of the k th element
P [k]	type number of the kth element used as a load
Q	type number of the variable for a given type of element
B	type number of the variable that is the parameter
B1	remembers value of B
t	number of output option
opt [t]	= 1 for option t, = 0 for no option t
h 1	total number of tees (series or shunt) encountered in series
h	number of particular tee (= 0, 1, 2 ..... h1)
m [h]	= 1 for shunt tee, = 2 for series tee
j	= 2 h.

APPENDIX 5

List of Network Elements

The values of all parameters to be in M.K.S. units.

Type No. 1	Function	Parameters
1	shunt tee	
2	series tee	
3	branch load	network type 6, 8 or 10
4	final load	network type 6, 8 or 10
5	line	Zo, l
6	shunt R	Rs
7	series R	Rs
8	shunt C	Cs
9	series C	Cs
10	shunt L	Ls
11	series L	Ls
12	shunt T.D.	R, C, r, L, Cl (R is a positive number)
13	series T.D.	R, C, r, L, Cl
14	shunt s.c. stub	Zo, l
15	series s.c. stub	Zo, l
16	shunt o.c. stub	Zo, l
17	series o.c. stub	Zo, l
18	shunt table)	
19	series table)	spare
20	shunt element)	
21	series element)	

APPENDIX 6

List of Output Options

Option Number	Output	Symbol in Programme
1	Real and imaginary parts of input impedance	reZin, imZin
2	Normalized real and imaginary parts of input impedance	renZin, imnZin
3	Real and imaginary parts of input admittance	reYin, imYin
4	Normalized real and imaginary parts of input admittance	renYin, imnYin
5	Real and imaginary parts of reference characteristic impedance	reZc, imZc
6	Modulus of input reflection coefficient	modgamma
7	Angle of input reflection coefficient	arggamma
8	Available reflection gain in dB	ref Gav
9	Real and imaginary parts of tunnel diode admittance	reYD, imYD
10	Insertion loss in dB	dBins

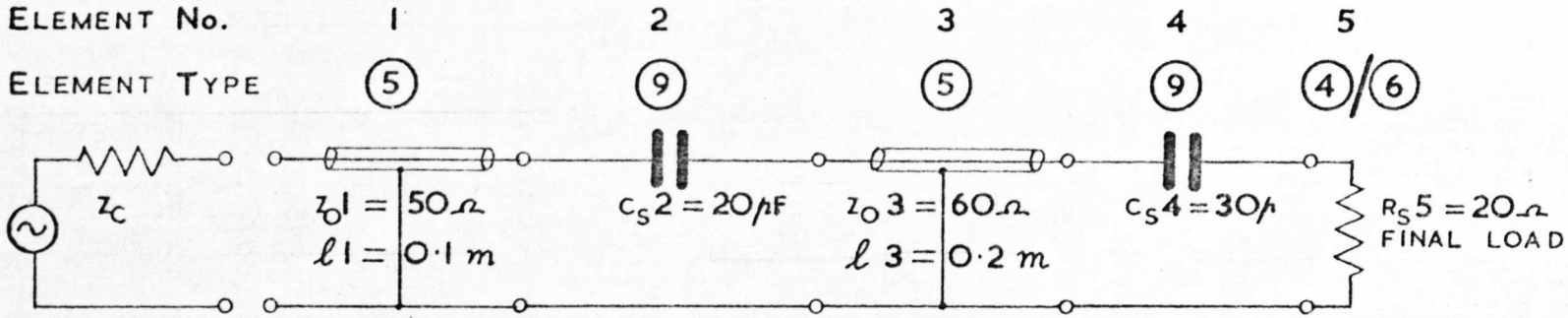


FIG. 1. EXAMPLE OF CIRCUIT WITHOUT TEES

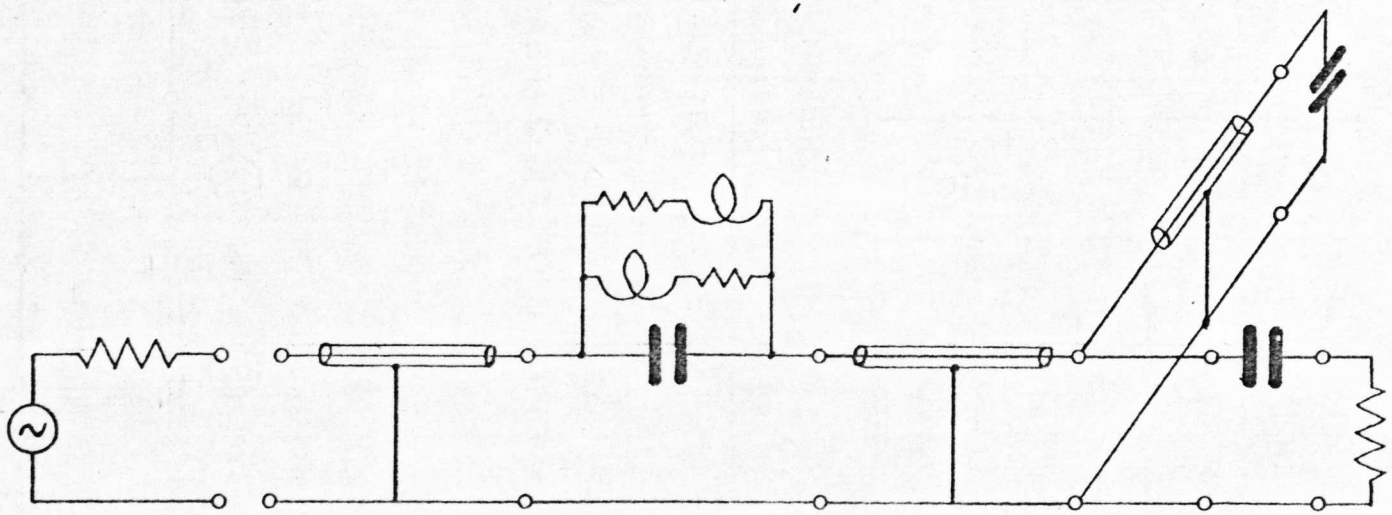


FIG. 2. CIRCUIT REQUIRING THE USE OF TEES

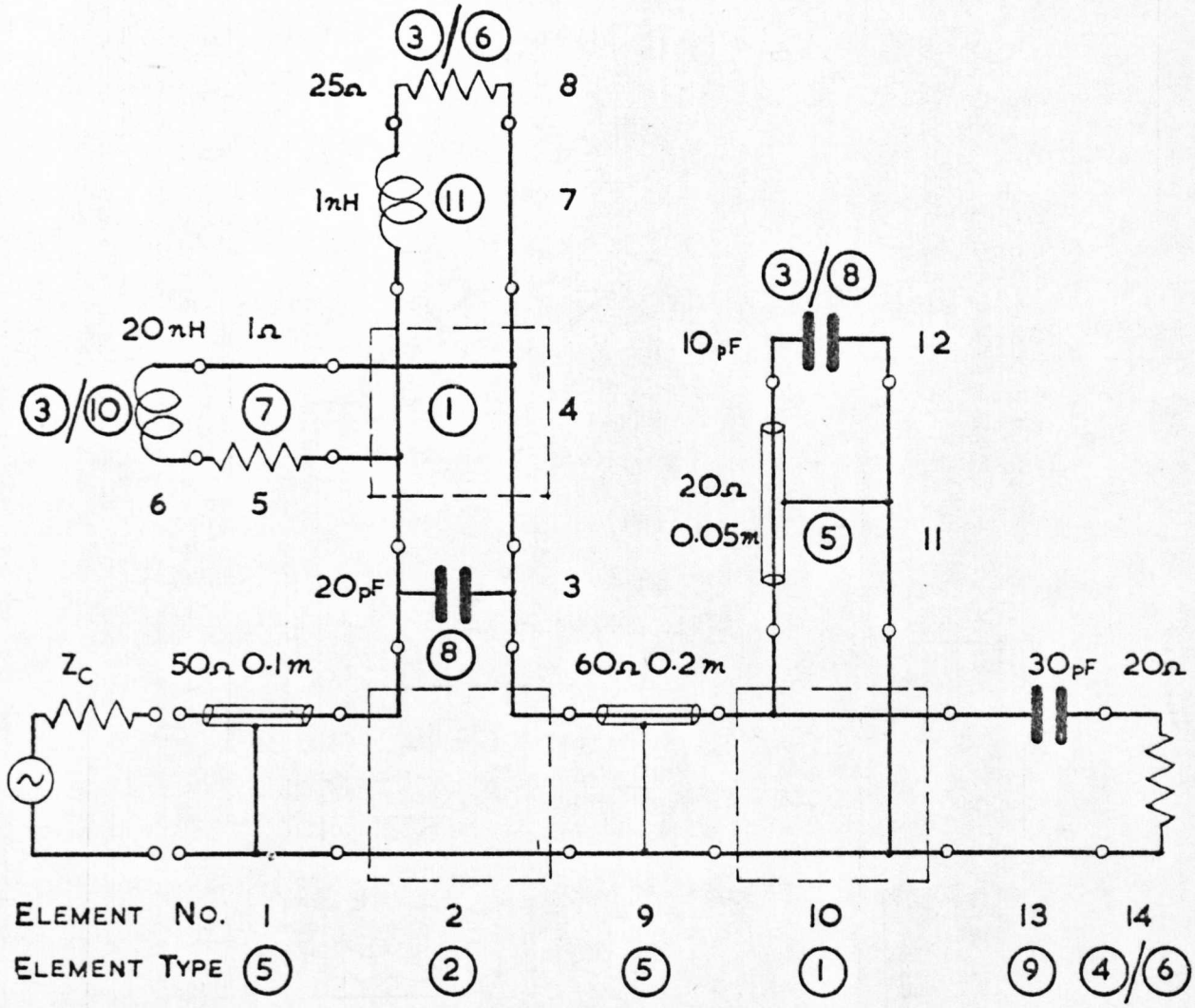


FIG.3. CIRCUIT OF FIG.2 REARRANGED USING TEES

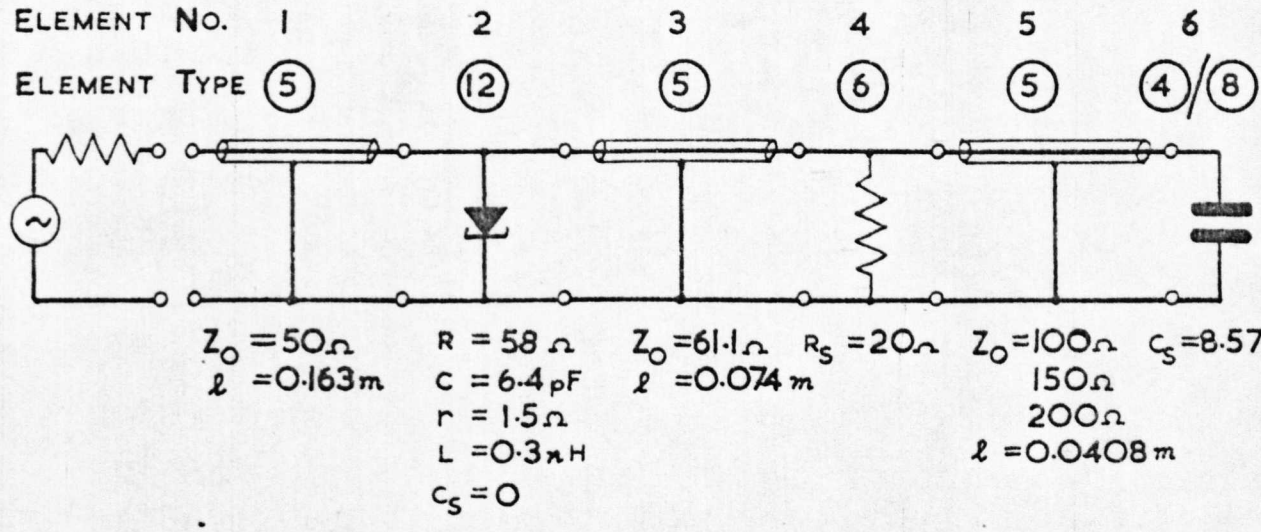


FIG.4. TUNNEL DIODE REFLECTION AMPLIFIER USED FOR EXAMPLE

```
1
£Example - tunnel diode reflection amplifier?
6
50
1  6  7  8  0
1     5     50   .163
2     12    58   6.4₁₀-12   1.5   .3₁₀-9   0
3     5     61.1 .074
4     6     20
5     5     ₁₀70  100  50  201  .0408
6     4     8  8.57₁₀-12

.3  .2  .71
```

Fig. 5. Example data tape for circuit of Fig. 4.

Example ~ tunnel diode reflection amplifier
n= 6 Zc= 50.00ohms
Output options 1 6 7 8 0

1 5 line
Zo= 50.000ohms
l= .1630m
2 12 shunt T.D.
R= 58.00ohms
C= 6.400_{10}-12F
r= 1.500ohms
L= 3.000_{10}-10H
Cs= 0.000_{10}+00F
3 5 line
Zo= 61.100ohms
l= .0740m
4 6 shunt R
Rs= 20.00ohms
5 5 line
Zo= 100.00 5.000_{10}+01 2.010_{10}+02ohms
l= .0408m
6 4 final load 8 shunt C
Cs= 8.570_{10}-12F

Zo 5= 100.0ohms

Freq (Gc/s)	reZin (ohms)	imZin (ohms)	modgam	arggam (deg)	refGav (dB)
.3000	133.8	-112.1	.6502	-21.84	-3.740
.5000	-46.86	22.16	4.439	85.17	12.95
.7000	-114.0	-101.3	1.609	-26.01	4.131

Zo 5= 150.0ohms

Freq (Gc/s)	reZin (ohms)	imZin (ohms)	modgam	arggam (deg)	refGav (dB)
.3000	361.2	-204.7	.8109	-6.871	-1.820
.5000	-36.53	31.07	2.715	93.69	8.676
.7000	-125.7	-106.1	1.575	-23.36	3.943

Zo 5= 200.0ohms

Freq (Gc/s)	reZin (ohms)	imZin (ohms)	modgam	arggam (deg)	refGav (dB)
.3000	-350.4	46.31	1.326	2.167	2.452
.5000	-31.01	31.31	2.372	100.1	7.503
.7000	-131.8	-106.7	1.568	-22.12	3.906

Fig. 6. Example results for circuit of Fig. 4.

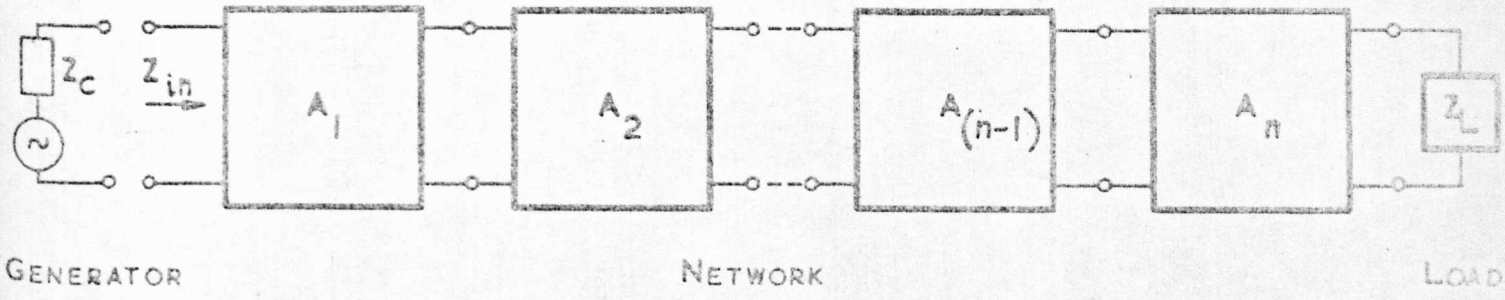


FIG. 7. GENERAL CASCADED NETWORK

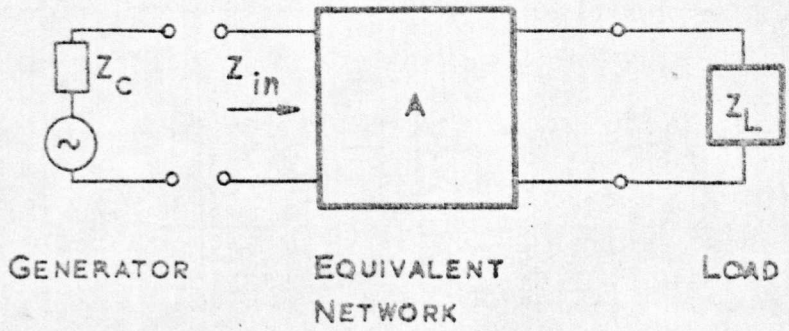
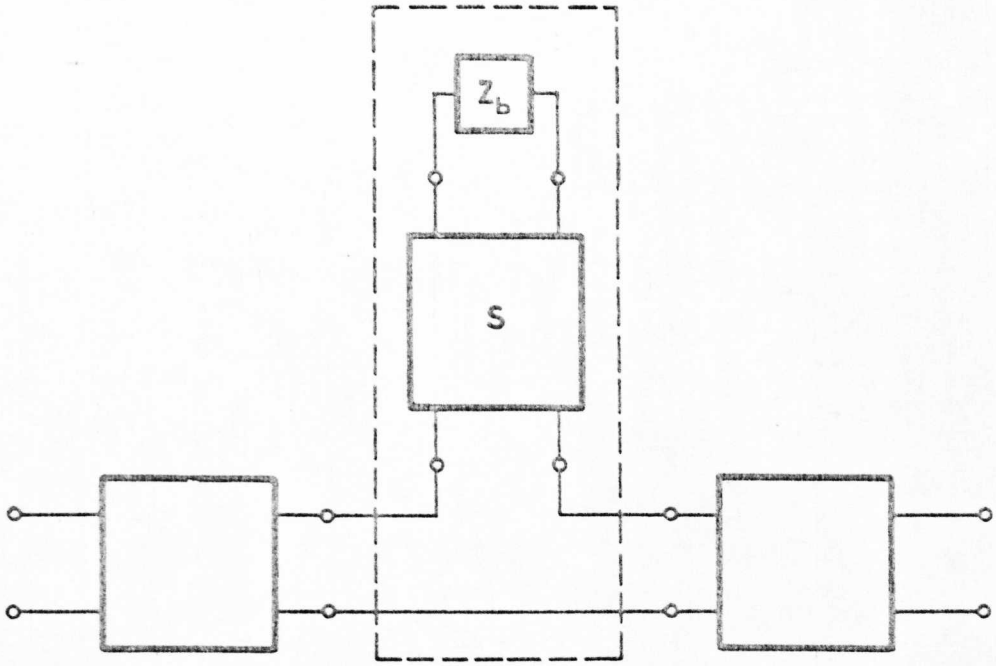
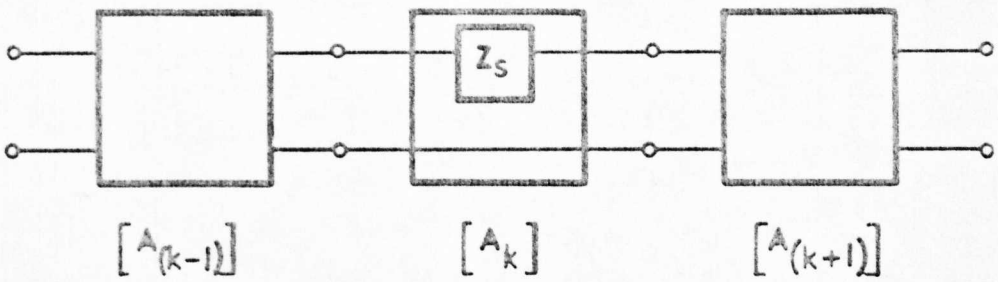


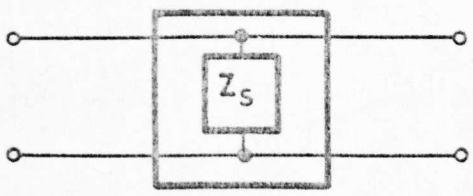
FIG. 8. EQUIVALENT NETWORK



(a) PART OF CASCADE CONTAINING A SERIES BRANCH CIRCUIT



(b) SERIES BRANCH REPLACED BY EQUIVALENT SERIES IMPEDANCE



(c) SHUNT IMPEDANCE EQUIVALENT OF SHUNT BRANCH

FIG.9. TREATMENT OF BRANCHES

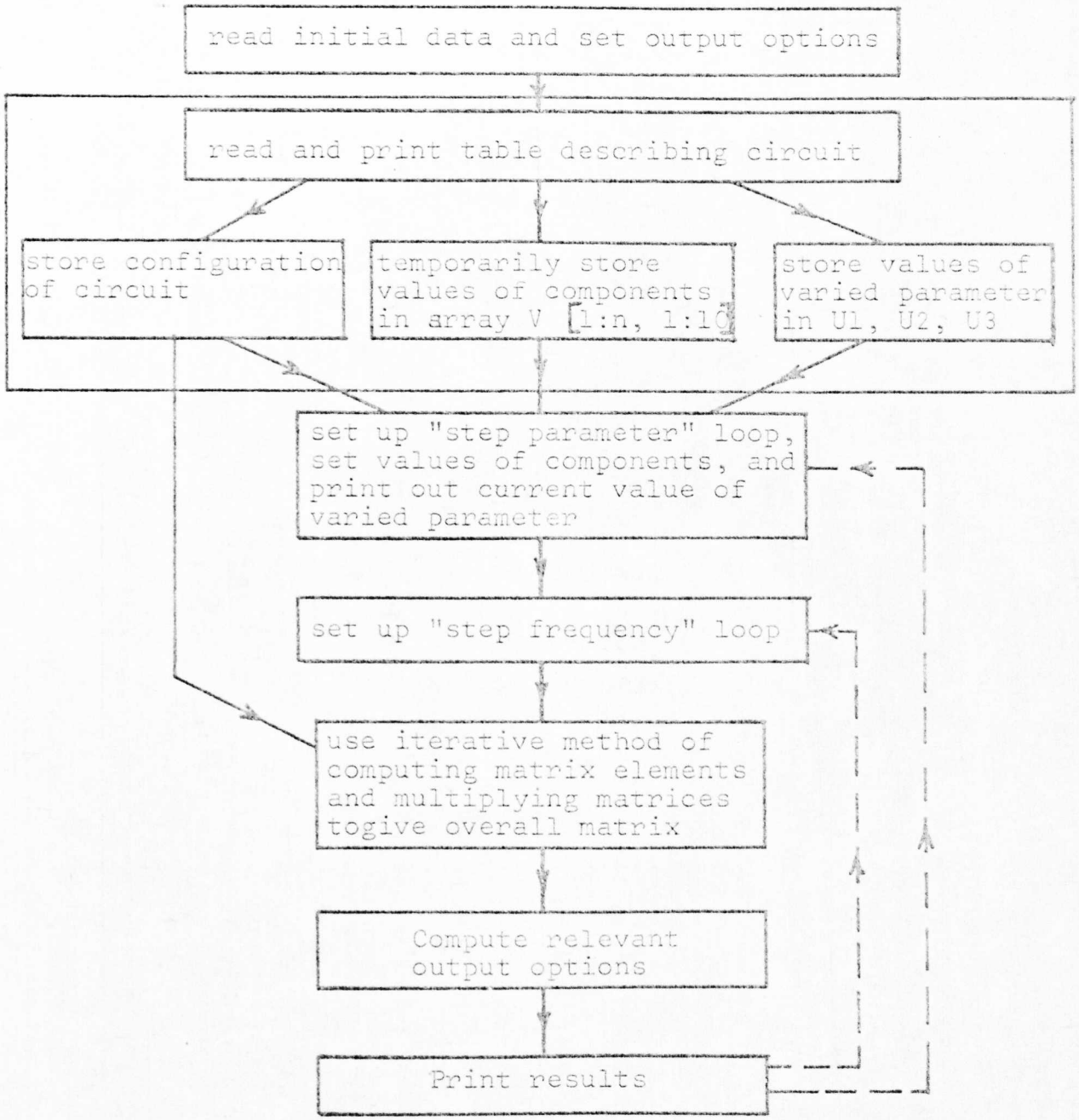


Figure 10. Flow diagram showing main features of programme

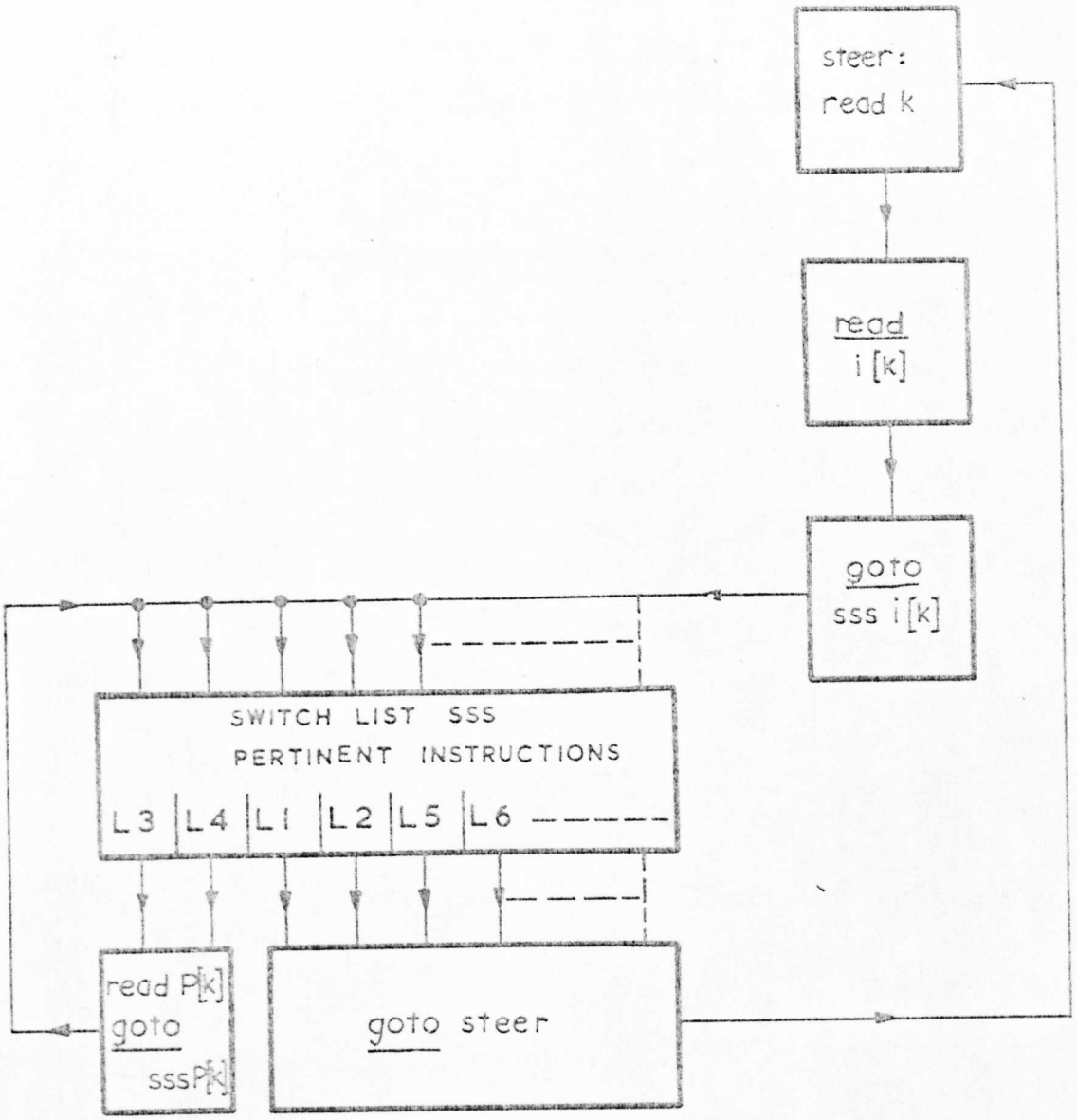


FIG. 11. READING AND STORING CIRCUIT CONFIGURATION

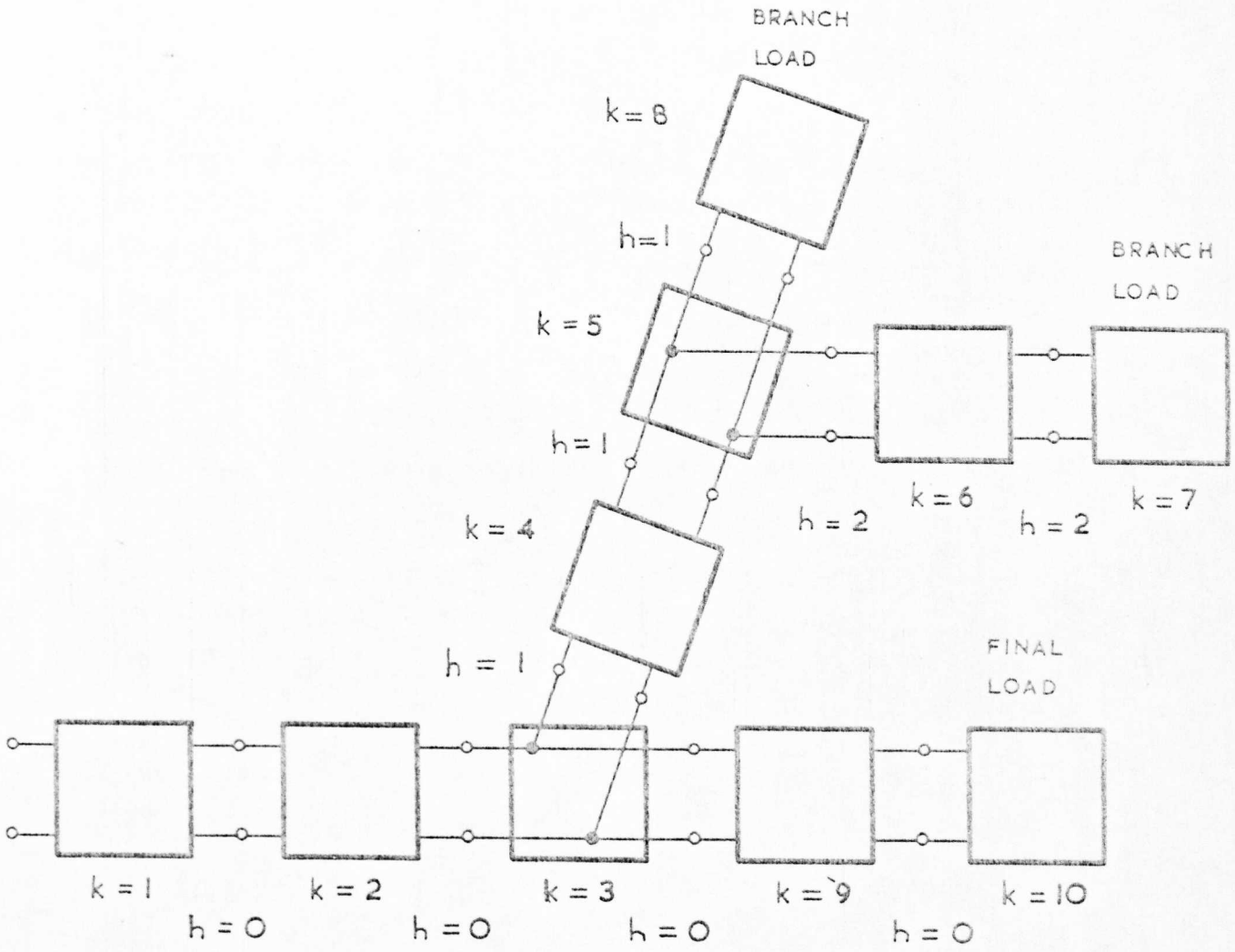


FIG. 12. KEEPING TRACK OF BRANCHES

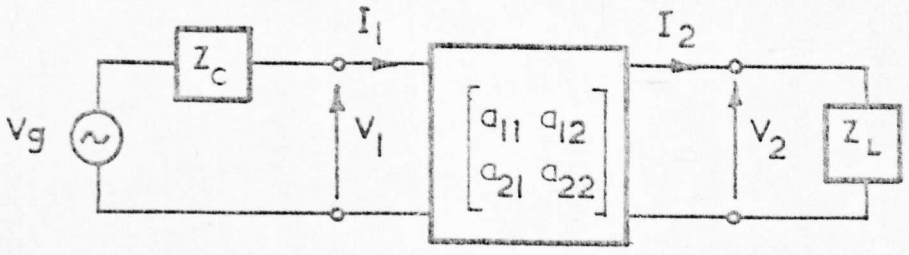
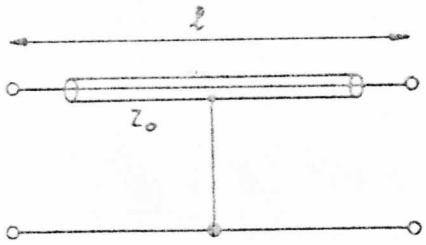
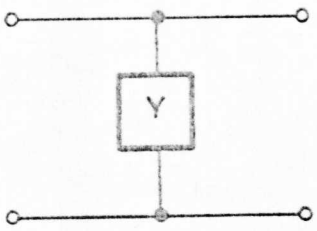


FIG. 13. CALCULATION OF INSERTION LOSS FROM CHAIN MATRIX



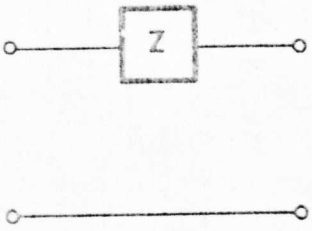
$$\begin{bmatrix} \cos\beta\ell & jZ_o \sin\beta\ell \\ \frac{j}{Z_o}\sin\beta\ell & \cos\beta\ell \end{bmatrix}$$

(a) LOSSLESS TRANSMISSION LINE



$$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$$

(b) SHUNT IMPEDANCE



$$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$$

(c) SERIES IMPEDANCE

FIG. 14. CHAIN MATRICES FOR CIRCUIT ELEMENTS

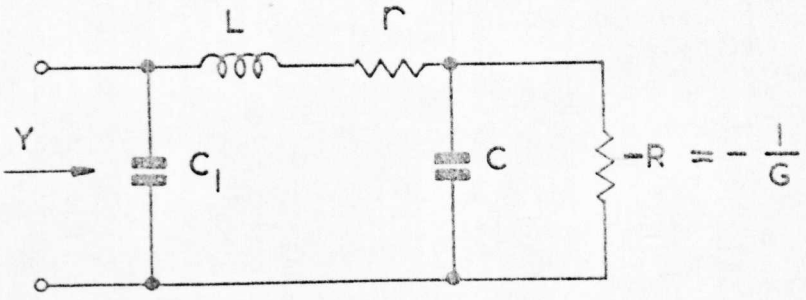


FIG.15. EQUIVALENT CIRCUIT OF TUNNEL DIODE

APPENDIX 15.

Automatic general-purpose microwave circuit analysis programs.

An invited contribution to a panel discussion.
I.E.E.E. Trans. MTT-17, 8, pp527-533, (Aug. 1969).

Automatic General-Purpose Microwave Circuit Analysis Programs

Panel Discussion

PETER E. GREEN, MEMBER, IEEE, MICHAEL K. MC PHUN, MARCO A. MURRAY-LASSO, MEMBER, IEEE, AND ALLEN E. SMOLL, SENIOR MEMBER, IEEE

THE PANEL members were asked to respond to the Editor's question, given below. They were then given copies of each other's comments, and asked to respond with a second-round of opinions.

Editor's Question

An automatic, general-purpose, circuit analysis program is one that accepts from the engineer a description of an arbitrary circuit configuration, with element values, and then sets up and solves the necessary equations automatically, to present the engineer with requested circuit performance data.

Programs of this type (such as ECAP) are widely used by engineers working at submicrowave frequencies, but because of their limitation to lumped elements, they are seldom used by microwave engineers.

However, automatic general-purpose circuit analysis programs suitable for microwave circuits have been developed to varying degrees of sophistication by individual, computer-oriented, microwave engineers, but very little has been published.

You have had direct experience with an automatic, general-purpose microwave circuit analysis program. What is your opinion of them; what future do you see for them?

First Round Comments

P. E. Green

These programs are becoming a very important tool of the microwave engineer. We implemented our company's first automatic general-purpose microwave circuit analysis program almost one year ago. This program is now used on a time-shared computer, for several hours each day by microwave engineers. It has become a tool for checking designs before they are built; further, we find engineers refining their designs on a trial and error basis using the program. From our experience with this and later versions I conclude that these programs are very worthwhile and are here to stay.

Our existing general purpose programs are limited to frequency domain analysis of microwave circuits. Most of these programs use transfer-matrix analysis techniques which were simple to implement. Although this works well for circuits with a tree structured topology, circuits with multiport elements cannot be analyzed by these techniques. Here, a nodal analysis program is used, the circuit being represented by an admittance or impedance matrix and reduced to a transfer matrix between the input and output ports. This technique, while allowing analysis of any linear microwave device, consumes large amounts of computer time and core, and is expensive to implement and use. The next step in the development of these programs is to combine these two analysis techniques; using transfer matrix techniques wherever possible. Then the complexity of circuits that can be analyzed will be limited only by the size of the computer available.

Nonlinear devices present a real challenge. At low frequencies the time dependencies of the circuit currents and voltages are calculated by numerical integration. Only lumped components can be used, but nonlinearities of diodes and transistors can be modeled accurately. At microwave frequencies, components are distributed in nature and must be treated as such for accurate results. Distributed components can be represented by a large number of lumped components. However, because of time-constant problems in numerical analysis, this results in long run times on large, expensive computers to analyze even simple devices.

Techniques are now needed to considerably speed the running of time domain analyses on distributed nonlinear devices. Existing techniques solve for the fields within these devices. For circuit analysis, only external voltages are of interest, and by eliminating the solution of internal fields the analysis process should be considerably shortened. Once these techniques are developed, a general purpose time domain analysis program for microwave frequencies will be feasible.

While good analysis techniques are essential to these automatic analysis programs, a very important, and often neglected problem, is the on-line interface with the engineer. A program must be easy to use to be a convenient and efficient tool. It must also be available when required and it must present the data in a useful form.

We have experimented with various forms of input language for these programs. Numerical codes to represent the circuits and analysis required, while easy to program, are

P. E. Green is with Sanders Associates, Inc., Nashua, N. H.

M. K. McPhun is with the University of Warwick, Warwick, England.

M. A. Murray-Lasso is with the Systems Research Center, Case Western Reserve University, Cleveland, Ohio.

A. E. Smoll is with Philco-Ford Development Laboratories, Palo Alto, Calif. 94303.

very unpopular with engineers as they increase the possibility of careless errors. Alternately, a completely English language form for input, while useful for training, is too slow for everyday use and quickly becomes annoying. We have found the most popular form of input language to be one using meaningful mnemonic abbreviations.

We have made these programs available on time-sharing, batch processing, and small dedicated computers. Each has its advantages but the most popular by far is the time-shared version. Flexibility and availability are the desirable characteristics here. Flexibility in that the engineer can input a circuit, analyze it, change it, analyze it again, all while seated at a teletype. Because it is on a time-shared service, it is always available at short notice, enabling an engineer to analyze a design without breaking his train of thought. This version has the disadvantages of a limited analysis capability and poor presentation of results. Analyzing a complicated circuit requires a large computer, presently available to us only on a batch processing basis. However, large time-sharing computers are now becoming available and I think that all microwave analysis programs will ultimately use time-shared computers.

The disadvantage of the time-shared programs is the presentation of results in tabular form on a teletype which have to be subsequently hand-plotted. One of the best forms of data presentation for a permanent record is the graphs produced by Stinehelfer's MACAP program. These are large 8×10 prints and are excellent for subsequent analysis of results. Experience with a CRT display driven from a small dedicated computer indicates that this is an excellent medium for trial and error computer-aided design. While a 6 inch screen is hardly the ideal medium, it does point the way to the use of large screen displays. Currently, these are luxury items, too expensive for everyday engineering use, but their cost is rapidly decreasing.

I conclude that the future microwave analysis programs will use a time-share computer, and they will present results for immediate viewing on a large screen display, with the option for producing a hard copy plot for later analysis. Data and the flow of the analysis will be fed in by means of a teletype keyboard using an abbreviated English language format. We have general purpose programs for the frequency domain analysis of microwave circuits. These must be refined to reduce the running costs and improve the interface with the engineer. I hope someday soon we will have time domain analysis programs the equal of CIRCUS or SCEPTRE for use with microwave circuits.

M. K. McPhun

Imagine you had at your disposal a broad-band microwave test bench capable of producing errorless frequency responses of complex impedance, insertion loss, etc., over any range of frequency you demand. Further, suppose your model shop could produce changes in your circuit design within seconds, what microwave engineer would turn his back on these facilities? A fanciful pipe dream you say, and yet these programs do just these things.

I believe that one of these programs is an essential tool of the microwave circuit designer. Using one, you can save a great deal of workshop and experimental effort. And if this were true in the past, it is even more so in these days of integrated circuit techniques. It is rare that we have an accurate theoretical model at hand for every aspect of a microwave circuit design. We must estimate a little or a lot, and then verify our design in practice. Sure enough, we must make something to test, but it does not have to incorporate a number of variable elements, we can vary these on the computer.

As an example, suppose that we have designed a simple broad-band filter/matching network consisting of a few lengths of transmission lines, and we have some idea of the stray capacities at the junctions. Before we make any hardware at all we can compute the broad-band responses we are interested in to ensure that the strays do not throw up some spurious peaks. (Even without the strays it is comforting to have the computer check that we did our sums right!) Having made the hardware we might want to investigate some deviation from the expected results. So we return to the computer and using our intuition we modify our theoretical circuit until the computed response matches the experimental results.

Before considering the future I will outline some desirable features that I feel should be incorporated in these programs.

1) An obvious point, but one that cannot be overemphasized is that the printout of results should contain a description of the circuit computed. Large quantities of results are not of much use if in a few weeks time you cannot identify the circuit for which they were computed.
2) It should be easy to write the input data for the circuit configuration, component values, and results required. If it is not easy, only those engineers who are computer-oriented will use it.
3) It should be possible to vary at least one parameter to obtain a set of responses without having to reinput the data. One parameter as a variable at a time is probably enough, otherwise it is only too easy to overburden yourself with results.
4) Automatic plotting of results by the computer on a digital plotter is a great help, but this leads us into conflict with 2) above. There are many forms that the plotted data can take, particularly when variable parameters are involved, and any of these forms might be required by the user. Quite apart from the additional complications involved in writing the analysis program, the variety of choice of output data formats would be confusing to the uninitiated user. My solution is to have two quite separate programs, one with plotting facilities and one without.
5) The design of the program must strike a compromise between versatility and other factors. For example, my program[1] will not handle feedback loops. Whereas this prevents its use for a few circuits, the acceptance of this

[1] M. K. McPhun, "A computer programme for the analysis of branched distributed and lumped circuits," *Electrical Networks*, IEE (London), Conf. Publ. 23, pp. 89–124.

limitation gives higher computing speed, and simpler input of circuit data.

6) Normally the user must decide what frequencies he wishes to compute. Too few frequencies and you may miss out a spurious resonance; too many and you waste money on unnecessary computer time and get overburdened in paper. I have devised and tested a routine that has the computer decide just how many frequencies to calculate to avoid missing our important data, but so far I have not incorporated it into the automatic analysis program.

We cannot consider the future of these automatic analysis programs without also considering the automatic synthesis program. We may ask, if we have programs which will synthesize a microwave circuit to have a given frequency response etc., is there then any need for the automatic analysis program? Of course, the analysis program will in general form the heart of the synthesis program anyway, so this is partly its place in the future. However, it is impossible to build the intuition of an engineer into a general purpose synthesis program. I believe the future lies in the engineer himself forming part of the synthesis process, using the analysis program in a conversational mode with the computer. To illustrate this we will look again at the example of the third paragraph.

At the end of the example, when the engineer is engaged in a synthesis process, he is in fact forming part of a hill-climbing routine. The snag here though is the time taken to prepare input data, wait his turn to get results from the computer, prepare new input data, etc. A personal terminal in a multiaccess system would be a great help here, but I feel that the real answer is the use of a CRT and lightpen system.

Let us then picture our engineer sitting in front of a CRT display. On the display he can call up a picture of the circuit diagram, tabulated results, or a wide variety of frequency responses. Most important, he can change component values at will (probably using the keyboard), or change the circuit diagram using the lightpen. When by repeated use of the analysis program he has achieved his ends, he can have the computer print out and plot data and results in a permanent form.

I am not ruling out the inclusion of the automatic synthesis program as part of this process, but where it is used it must be firmly under the control of the engineer. Also the synthesis program would normally be restricted to optimising component values, whereas the engineer can make changes to the circuit configuration.

A word of warning—these programs can make us lazy. In some circumstances there is still no substitute for a little algebra to find an optimum solution.

M. A. Murray-Lasso

In order to discuss automatic general-purpose microwave circuit analysis programs, the particular needs of microwave engineers must be considered. The area of microwave circuits is a branch of electromagnetic field theory and as such, the principal physical law involved is Maxwell's equations. Although Maxwell's equations in empty space are linear, because they are partial differential equations, their solutions depend heavily on the boundary conditions. The enormous variety of boundary conditions possible virtually eliminates the hope of obtaining simple analytical solutions. Either numerical methods are used or complicated infinite series result even in simple cases. The question is more complicated for fields in matter where nonlinearities and inhomogeneity can occur. In practice, however, most microwave engineering work deals with linear systems.

Luckily in the practice of electrical art, simplifying assumptions can be made which for slowly varying fields give excellent engineering results.[2] The most useful one consists of assuming conservative fields, concentrating the analysis at discrete spatial positions and using only some zero and first order terms in the Neumann series (quasi static approximation) of the solution of Maxwell's equations. In particular, coupling between zero order electric and magnetic fields gives rise to "resistors," coupling between time varying zero order electric fields and first order magnetic fields give rise to "capacitors," and coupling between time varying zero order magnetic fields and first order electric fields give rise to "inductors." On these simplifications is based the usefulness of circuit theory to describe physical equipment for low frequencies.

At microwave frequencies, a little more sophistication is required for good correlation between circuit theory and field theory, hence, distributed circuit theory is used for many devices where the frequency dependency can no longer be approximated by ratios of polynomials of low order. When the fields are such that certain assumptions (such as TEM modes) are no longer valid, combinations of theoretical, experimental and intuitive techniques are used to describe the behavior of devices. Nevertheless, the methods usually retain a strong circuit flavor.

Because of the type of situations that microwave engineers face, a general purpose computer program to analyze microwave circuits must have characteristics which differ in some aspects from those of electronic circuit analysis which are in wide use presently. Since most of the work in microwave circuits involves distributed linear time invariant systems, frequency domain methods are very useful, in contrast to the situation in electronic circuits where much of the interest is in lumped nonlinear circuits. The state-variable-first-order-vector-differential equation approach is less useful in microwave circuits. There is also interest in transients for microwave circuits. These may be computed efficiently using fast Fourier transform algorithms, which recently have received considerable attention.

As mentioned above, it is difficult to model microwave components accurately with conventional circuit elements (R, L, C, M) without introducing a great number of them. Since the number of elements a program can accept is limited by computer memory, one is forced to resort to more convenient modeling. The first simplification that can

[2] These engineering approximations were known before Maxwell's equations were discovered.

be made is to use analytic expressions for frequency domain characterization of certain standard devices like coaxial lines, RC lines, directional couplers, waveguides, etc. These usually require only a few electrical parameters per unit length plus dimensions. Additional methods must be provided to accommodate elements whose terminal characteristics have been determined experimentally or calculated at discrete frequencies using other special purpose programs. For this purpose, it is desirable to be able to enter circuits described by tables or through formal expansions in terms of standard functions (such as polynomials).

An additional feature very desirable for microwave circuits is the possibility of attacking large circuits by preanalyzing small pieces, connecting the pieces and then analyzing the whole circuit. In this way, complicated stray effects can be modeled without exhausting the memory of the computer.

The types of input and output which a program for microwave networks should handle often differs from that of an electronic circuit program. For example, at high frequencies, scattering parameters are much easier to measure than impedance or admittance. The engineer should be able to input multiterminal circuit data in the form in which he has measured it without going through laborious conversions. This should be done directly by the computer. In many circuits certain items appear several times in the same circuit. (For example, directional couplers, stubs, etc.) The user should be required to input the description of such devices only once, and only indicate where the same device appears. The program should use the analytical expressions or tables several times without unnecessarily overloading the computer memory.

The output that a microwave engineer wants is usually presented in a different form from what an electronic designer wishes. In microwave circuits engineering quantities like return loss, insertion loss, standing wave ratio and reflected power, are of interest while an electronic designer is interested for example in currents flowing in the branches, a situation which seldom is of interest in microwave circuits. It is highly desirable to obtain output in graphic form for ease of performance evaluation.

Microwave circuits are usually more sparsely connected than electronic circuits. In calculating a transient by fast Fourier transform, one may require several hundred frequency points. Thus, one should exploit the sparsity of the circuit as much as possible. This can be done by ordering the equations properly. It pays to spend some effort to order the equations if one is going to solve the same circuit several hundred times. Other methods like partitioning and iterative schemes also save much computation in frequency response solutions.

The end objective of circuit analysis is generally to aid a designer to determine parameter values which insure proper performance of a circuit. This means that a good program will be much more than an analysis program. It should have a library of devices. It should have additional options like automatic parameter variations, sensitivity, statistical variability, and worse case. It should also have selective printed and plotted output in convenient formats. All these increase the utility of a program when used to design.

Finally, through the incorporation of automatic parameter optimization (or in rare cases direct synthesis algorithms) a program can directly provide a designer with the final values of unknown parameters which make the circuit perform in a desired manner. This saves the designer the effort of repeatedly approaching the computer, analyzing results and changing parameters until his design is complete. This topic will be treated in more detail in a paper in the body of the issue.

In short a microwave circuit analysis program should provide an engineer with a convenient "laboratory" in which he can test ideas as much as possible without having to resort to building the device until he is almost certain the device will perform as desired. Some programs incorporating some of the features mentioned above are already in existence. More work to improve them is desirable. When good programs are available it will be too expensive for designers not to use them.

A. E. Smoll

As in many other engineering fields, the computer is having a profound effect on microwave design and development. Its penetration as yet has not been as great in microwaves as, for example, in lumped-circuit networks, probably because of the microwave engineer's heavy dependence on empirical data. Nevertheless, several organizations are successfully employing computer programs for general microwave circuit analysis. It may be interesting to describe briefly our experience with general purpose microwave circuit analysis programs.

Several years ago we developed a program to run on the Autonetics RECOMP II computer, which is a desk size machine with a 4098 word (48 bit) memory. The engineer conversed with the computer, i.e., entered his commands and then received his answers, through the on-line flexiwriter. The analysis proceeded in the same manner that he would make calculations on a Smith chart. To move to a new position along the guide or to add an immittance, he would type in the appropriate command code along with the numeric data. Thus, the circuit would be described by a sequence of such operations. Options for the calculation and typeout of various quantities such as VSWR and transmission loss were available at any point in the circuit. This program was very useful in designing a circuit to match an arbitrary impedance, such as for an antenna, and for obtaining the actual frequency response of filters. Because of the relative limited capacity of the RECOMP computer, this early program had a fairly simple command repertoire.

A somewhat more sophisticated program has been coded recently in FORTRAN for the GE Mark II time-sharing system. However, the high cost encountered during the checkout and operation of the program has discouraged its use. The overall speed is roughly the same as that of the RECOMP program since both converse through a typewriter. The use of a general-purpose microwave circuit program through a typewriter on the present time-share computing systems

would be economically feasible only if it must be used frequently for urgent purposes and if the amount of output is limited. Our program consisted of approximately 24 000 characters, too long to read for each usage.

A more favorable throughput cost has been attained with batch processing on a high speed digital computer with program requirement of less than 32 000 words of storage. The input consists of the program card deck followed by data cards representing the cases to be analyzed. Not only can one gain more calculation and output per dollar, but generally less of his time is required. The program is much more sophisticated in the command repertoire and the output options. Also, other output options such as plotting are available, and protection is provided against many types of inadvertent user errors. The drawback with batch processing is the long waiting time between submission of the job and receipt of the output. However, the turn-around time may improve dramatically soon for the generally available computing services, thus accentuating the use of large, sophisticated, general-purpose microwave analysis programs.

Only the analysis of the general class of circuits constructed by the successive cascading of two-port elements, i.e., circuits having one signal path, has been discussed here. This class includes the majority of applications. The preparation of a general-purpose analysis program for such circuits can be accomplished easily by a microwave engineer with the aid of a high level programming language such as FORTRAN. Programs for analyzing the general microwave network consisting of multiple-port elements interconnected in a general manner would be more complicated and probably useful for only special cases (for example, multimode waveguide elements). In designing a general-purpose program, the engineer must delineate the basic elements ("building blocks") of the circuits he expects to encounter and determine the mathematical transformation produced when a basic element is added to the preceding elements. Transformation by means of the transmission line equations or a matrix representation seem to be equally good choices. The matrix representation is somewhat more general and lends a flexibility for easier handling of special situations, e.g., symmetrical configurations. The appropriate mathematical model must be found for the discontinuity element, e.g., iris, bends. An equivalent circuit whose branches are evaluated from theoretical equations or empirical data may be used for this purpose.

In the author's opinion the general-purpose microwave analysis programs will have a prime role in microwave design and development. Many microwave engineers are going to acquire this tool and use it to replace much of the "cut and try" design on the bench.

It is interesting to speculate what effect the computer will have on the microwave engineering profession and technology. The computer will most likely upgrade the microwave professional environment. A greater interest will be created in more sophisticated microwave devices, ones that previously were unfeasible to consider developing. Increased emphasis will be placed on mathematical analyses and modeling of microwave elements.

Second Round Comments

P. E. Green

I think that the comment by M. A. Murray-Lasso about the inclusion of the results from computer controlled measurements is important. We have been operating a computer controlled microwave network analyzer for approximately nine months and have found it very useful in the development of microwave components. It will measure and print out such characteristics as VSWR, transmission loss, and impedance at specified frequencies. Currently, the only interaction with the analysis programs is in the comparison of experimental and theoretical data. However, we are planning to dump the measured impedance or transfer matrix of a device on paper tape and then to use this as data in one of the analysis programs. This data will be stored on disk and be accessed by the program as required. Initially, we intend to use this for interactively designing antenna impedance matching networks.

There is much interest in optimization algorithms. However, to misquote M. K. McPhun, "ten cents worth of algebra is worth a thousand dollars of computer running wild." I have experimented with many of these algorithms and found that they are only worthwhile in finally refining a solution. Starting with blind guesses for parameters usually ends with solutions requiring negative line lengths and the like. On the other hand, interactive optimization between the engineer and a time-shared computer works well. One aid that could be useful here is to print out the sensitivity of VSWR, transmission loss, and other performance characteristics to changes in each of the circuit parameters. This would enable the engineer to see which parameters to vary in order to improve the performance.

A. E. Smoll's comment about building blocks that can be inserted as required is very interesting. The nodal analysis programs can analyze multiport devices. A multiport device could be defined to be built from elements and other devices. This could be nested to many levels to allow the description of very complicated circuits. Further, the maximum size of matrix that has to be inverted is the maximum number of nodes in a device, the analysis being performed on a device that is composed of other devices. This would allow these types of programs to be run on smaller machines, providing disk storage was available.

Finally, I disagree with Mr. McPhun, that these programs will make us lazy. Properly used, they will serve to make the engineer aware of the problems with his design before and not after he has built it.

M. K. McPhun

My first reaction on reading the other contributions was to agree with everything said, but obviously I was writing for a different reader than the others. Who are our readers? Microwave engineers, obviously. How much experience of general circuit analysis programs of any type have they? In Great Britain the answer is certainly very little, if any, but of course I cannot speak for the U.S.A. The work on developing more elegant and versatile programs will continue

regardless, but there is a need now for some vigorous "selling" of the programs that already exist, to secure their wider use. As I see it, this issue of the TRANSACTIONS is fulfilling a valuable function here.

My experience so far has been with batch processing. However, I have been lucky in that I could run three batches in a day. The experiences of M. A. Murray-Lasso and P. E. Green with time-sharing systems seem to have led them to quite different conclusions from each other.

I agree with A. E. Smoll that the interconnection of multiport elements in a *general* manner is not generally necessary for microwave problems, and such a facility in a program would be rarely used. Analysis using transfer matrices can be very versatile. An unlimited number of branches can be tackled this way, but loops are forbidden. Where three-port elements are encountered, the third port can be treated as a branch. Thus, multiport networks can be computed two ports at a time, providing there are no loops. I like P. E. Green's suggestion of combining the transfer matrix and nodal analysis techniques, using transfer matrices wherever possible. This should be very economical of computer time, as after reduction using transfer matrices, the number of nodes in the resulting network would be very small.

The attributes of a good program listed by M. A. Murray-Lasso are all highly desirable. However, for such a program to be of general use by the uninitiated the conversational mode of operation is essential. The same applies to my comments on graphical presentation. Thus, the initiative would be with the computer to ask the relevant questions at the right time. Thus, not only can the computer ask for the data needed, but also as the design proceeds it can tell the engineer about the more sophisticated options that exist. We have just taken delivery of a CRT display system, and intend to incorporate these features. In the meantime, versatile programs already exist, and microwave engineers are missing out by not using them.

M. A. Murray-Lasso

With respect to P. E. Green's comments on transients for nonlinear circuits with distributed elements, I feel the way to go is fast Fourier transforms rather than lumping and integrating ordinary differential equations in time. When one lumps, for example an RC line, the matrix associated with the set of differential equations should, as the number of lumps increases, have eigenvalues which tend to the eigenvalues of the distributed circuit. These, in the case of a shorted RC line, extend to infinity on the negative real axis. (In the case of a lossless line they extend to infinity on the imaginary axis.) The resulting equations are stiff. The large spread in eigenvalues gives considerable numerical trouble, forcing the integration time steps to be very small to insure numerical stability. The part of the response associated with the largest eigenvalue, contributes very little to the total response since it dies away very fast. In spite of this, the largest eigenvalue is the one that controls the integration time step and forces it to be small.

If we could remove the higher modes we would get rid of the numerical stability problem without affecting the answer significantly. This could be done by diagonalizing the A matrix in the state equations, and ignoring the modes of the largest eigenvalues, but it would be very expensive computationally.

On the other hand, using fast Fourier transform methods, the high frequencies are easily eliminated in a natural manner without paying any computational price. The numerical stability problem disappears completely. If an analytical expression is available for the frequency representation of the distributed portion of the circuit, the method of iterating in the time domain between the linear distributed part and the nonlinear static part is not inefficient because generally most of the time spent in fast Fourier transforming is used in evaluating the frequency points. The method has been tried by Silverberg of Bell Laboratories and Wing of Columbia University.

All panelists made comments on time-sharing and I heartily agree with them that there is where the field will go. In spite of the existence of automatic parameter optimization, we need to improve our knowledge on how to obtain parameters to initiate a search, how to adaptively change optimality criteria, and how to a priori express mathematically what we really want a circuit to do. To acquire this knowledge, there is the need to have the man in the loop. He is also necessary to make decisions at the higher levels in multilevel optimization, for example, changes in structure. For this, it is necessary to improve man–machine communications with better computer-aided design languages, where a man can direct computations from a terminal by being able to inspect selectively in graphic form results which aid him in taking decisions. A user should be able to store different versions of circuits in files for future additional comparison. He should be able to look selectively into the contents of a file and obtain a permanent record of it on command.

All these features cost money and hence, realistically, a compromise has to be made. The price of memory and graphic peripheral equipment is on a downward trend so it is possible to envision these items as economically feasible in the future. In the meanwhile time-sharing with relatively inexpensive graphic output is already here.

Finally I agree most emphatically with M. K. McPhun, a little algebra often does wonders.

A. E. Smoll

There should be no doubt after reading this issue that the computer is a prominent tool in the development and the design of microwave devices. Many engineers now possess the capability to analyze most microwave circuits routinely on a computer. With the ready availability of reasonably priced computational power it is a needless risk not to evaluate thoroughly a tentative circuit design before the hardware commitment.

The writing of a microwave circuit analysis program is relatively easy. However, several points are worth reemphasizing to those starting to prepare their own program. A prime decision must be made at the outset as to what type of a computer to use. Generally, the selection is between

1) a time-sharing system with a remote terminal where the input and output is by means of a keyboard and a one-character-at-a-time printer, and 2) a batch processor with off-line high speed input and output devices.

Usually, the computational costs and engineering time will be greater with the time-sharing system. However, the advantage of evaluating almost immediately one's tentative design can be tremendous. Most modern (batch process) computer centers today can only offer to provide the results a half day or later after the job has been submitted. Even so, the advantage of fuller and well formatted output can be worthwhile. Also, the dedicated computer can generally handle a much larger and more sophisticated program as well as provide plots and other types of output not possible on the typewriter terminal.

On programming itself, a few remarks may be worthwhile. It is better to use alphabetic mnemonic codes rather than an extensive table of numbers to describe the circuit elements (or operations). This has been the usual choice of those who have written and used circuit analysis programs.

Utility options can save computer and user time. For example, storing a portion of the circuit that is used several times can effect such savings. Another is an editing routine which allows the user to change a portion of the circuit without going through the whole list of elements.

Though mathematical formulas exist for many different circuit elements and thus, can be incorporated directly in the program, one is likely to meet situations where the element characteristics are not so formulated. Therefore, it is well to have the capability to enter tabular characteristics.

Mistakes and errors during the program usage seem unavoidable. The program should plan for them, for instance by checking the reasonableness of the input values or by reading alphabetic identifiers built into the data input. Recovery and/or escape procedures should be provided for the user after the mistake has been made.

Accurate and adequate documentation of a program cannot be overemphasized. Without good documentation the effort and money invested to create the program may be wasted.

APPENDIX 16.

A study of lumped components for microwave integrated circuits.

Abstract of a paper presented at the Symposium on Microwave Integrated Circuits, Nottingham University, September 1967.

SYMPOSIUM ON MICROWAVE INTEGRATED CIRCUITS

A Study of Lumped Components for Microwave Integrated Circuits

This paper investigates the possibility of using lumped elements rather than distributed elements for microwave integrated circuits. The advantages and limitations are discussed in some detail.

The suitability of various substrate and conductor materials is discussed together with manufacturing processes. It is shown that the use of dimensions and tolerances similar to those met in conventional silicon circuitry can yield a sufficiently wide range of component values for the production of microwave circuitry to 12 GHz.

The physical construction of lumped thin-film capacitors, inductors and resistors is explained.

There are special measurement problems presented by these minute components and several possible measurement techniques are discussed. A synopsis of the particular method used at M.R.L., including the use of a computer to facilitate rapid evaluation of results is given.

Results which have been obtained for both series and parallel resonant circuits from 5 - 10 GHz will be presented. Results will also be given for a network consisting of a parallel resonant circuit in series with a resistance: this can be used for stabilising a tunnel-diode amplifier.

Future developments include the use of different conductor materials to obtain higher Q-factors; an investigation into the microwave properties of thin-film dielectrics, and the integration of an unencapsulated diode chip with the thin-film circuitry to form, for example, a tunnel-diode amplifier.

M.K. McPhun,*
School of Eng. Science,
University of Warwick,
Coventry, Warwick,
England.

J.F. Wells,
Mullard Research Laboratories,
Redhill, Surrey,
England.

*formerly with Mullard Research Laboratories,
Redhill, Surrey, England.

APPENDIX 17.

Microwave measurement of thin film dielectric properties.

Conference on Dielectric Materials Measurements and Applications, Lancaster, July 1970, I.E.E. Conference Publication No. 67/70, pp 42-45.

MICROWAVE MEASUREMENT OF THIN FILM DIELECTRIC PROPERTIES

K. Mehmet and M.K. McPhun

In order to evaluate thin dielectric films for use at microwave frequencies it is essential to be able to measure their electrical properties. However it is very difficult to measure the dielectric constant and loss of a thin film because of its small volume. The only known published work (1) on microwave measurement of thin film dielectrics was at about 2.9GHz. For accurate results this technique required film thicknesses of order 10μm. Other published work (2) describes attempts to measure thin film dielectric properties using the film as the dielectric of a sandwich capacitor. Although these methods are simple, the measured properties cannot be calculated exactly because the series resistance of the electrodes cannot be determined accurately. Also we wish to determine what effect the electrodes have on the properties of the dielectric of a sandwich capacitor.

To overcome the above difficulties we are using the capacitively loaded coaxial cavity illustrated in Figure 1a. The cavity consists of two sections: the centre conductor of the top section is flush with the outer conductor, whilst the centre conductor of the bottom section is shortened by about 20μm. to give the capacitive gap, G_o, between the centre conductors. The mode of resonance used in our measurements is shown in Figure 1b.

In use, the thin film is deposited onto the end of the centre conductor of the top of the cavity. After completing measurements on that film it is then a simple matter to lap off the deposit and reuse the cavity for another measurement.

Theory: The problem has been treated using an equivalent circuit approach and the series tuned circuit configuration was chosen for simplicity. The cavity with no dielectric is represented by Figure 2a where C_o is the gap capacitance and mC_o the capacitance due to the fringing field. Figure 2b shows the cavity equivalent circuit with a dielectric film filling a fraction, x, of the gap. Solving these equivalent circuits, the relative dielectric constant of the film is:

$$\epsilon_r = x \left\{ \left[(m+1) \frac{f1}{f2} \frac{\operatorname{Tan} \frac{2\pi l}{v} f1}{\operatorname{Tan} \frac{2\pi l}{v} f2} - m \right]^{-1} + x - 1 \right\}^{-1} \quad \ldots.(1)$$

and the loss tangent of the film is:

$$\operatorname{Tan}\delta = \frac{\epsilon_r}{x} \left[\frac{S2}{S1} - 1 \right] \frac{\left(m \left(\frac{x}{\epsilon_r} - 1 + x \right) + 1 \right)^2}{(m+1)} K \quad \ldots.(2)$$

K. Mehmet and M.K. McPhun are with the School of Engineering Science, University of Warwick, Coventry.

Where m is fringing capacitance as a fraction of C_o
x is the fraction of gap occupied by the film
f_2 and f_1 are the resonant frequencies of the cavity with and without the film
21 is the length of the cavity
v is the velocity of light in free space ,3×10^8 m/s.
S_1 and S_2 are the voltage standing wave ratios (VSWR's) at the input to the cavity at resonance without and with the film
K is a constant which depends on the cavity material and dimensions.

Experimental Work: The cavity was made of silver plated Invar because of its low coefficient of expansion. Tarnishing of the silver plating was eliminated by plating with a thin layer (6 nm) of gold. The cavity was designed to resonate at about 9.5 GHz. in X-band and coupled to a waveguide through a dumbbell coupling aperture. Stringent measures were taken to ensure a rigid construction to give repeatable results upon repeated dismantling and assembling of the cavity. The cavity properties were measured using a precision reflectometer, consisting of high directivity waveguide directional couplers and a Hewlett Packard harmonic Frequency Converter and Network Analyser assembly, as the receiving system. Thus the phase as well as the magnitude of the reflection coefficient could be recorded as a function of frequency on an X- Y recorder. The arrangement is shown in Figure 3.

The measurement procedure is as follows. The cavity input reflection coefficient, V.S.W.R., resonant frequency and Q-factor are first measured. Then the top of the cavity is removed and the dielectric film is deposited onto the plane end of the inner conductor. The cavity is again assembled and the new resonant frequency and V.S.W.R. at resonance are measured for the dielectric loaded cavity. Using these measured results and equations (1) and (2), ϵ_r and $\tan\delta$ are determined.

For a typical cavity it was found that the measured unloaded Q-factor was 95% of the calculated value and the resonant frequency within 0.02% of the calculated value. So far we have deposited films of SiO_2 by evaporation using an electron beam evaporating source. The substrate (cavity) was maintained at 200^0C during evaporation. Measurements on these films of SiO_2 of 2μm. thickness gave ϵ_r between 3.9 and 4.2 and $\tan\delta$ of 0.008 to 0.014. It was observed that higher rates of evaporation gave lower values of $\tan\delta$.

Work is in progress to verify the accuracy of equations (1) and (2), and hence the method as a whole, by measurement of known material. This can only be done by mechanically and chemically thinning a section of a bulk dielectric of known properties until it can be inserted in the gap and measured by this method. At the time of writing we are also engaged on measurement of dielectric films of other materials by evaporation and R.F. sputtering. We hope to present the results of these experiments at the conference.

References:

1. Measurement of the Permittivity of Insulating films at Microwave frequencies. H. Sobol and J.J. Hughes, IEEE Trans. June 1967, MTT-15, pp. 377-378

2. Designing lumped elements into Microwave Amplifiers, M. Caulton and W.E. Poole, Electronics, April 14, 1969, pp. 100-110.

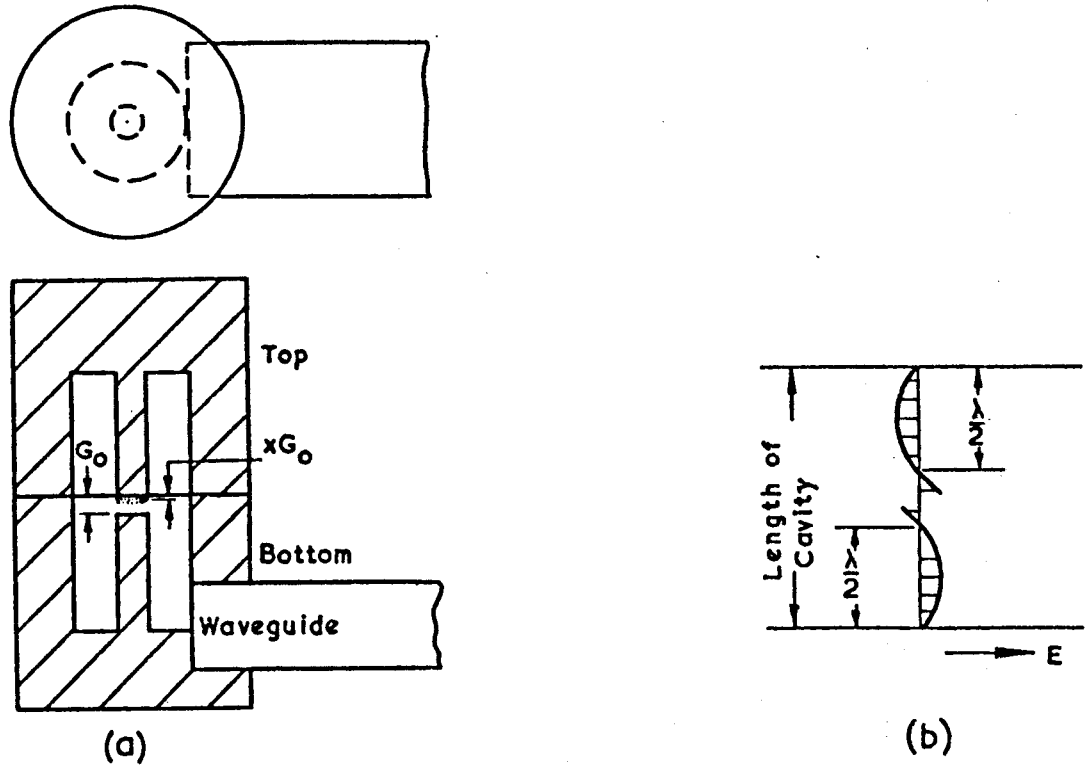


Figure 1a. Axial cross-section of cavity (gap greatly exaggerated)
b. Electric field distribution of cavity.

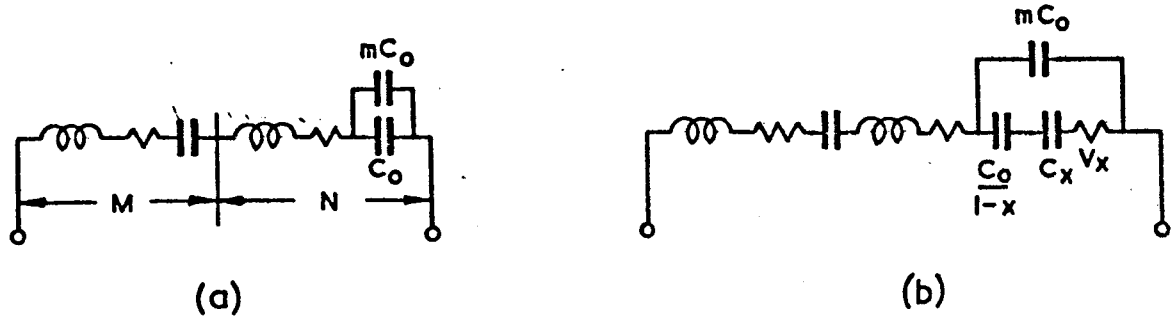


Figure 2. Equivalent Circuits, M represents the extra wavelength and N the length between the half wavelengths of the cavity. (a) Cavity with no dielectric film and, (b) Cavity with film.

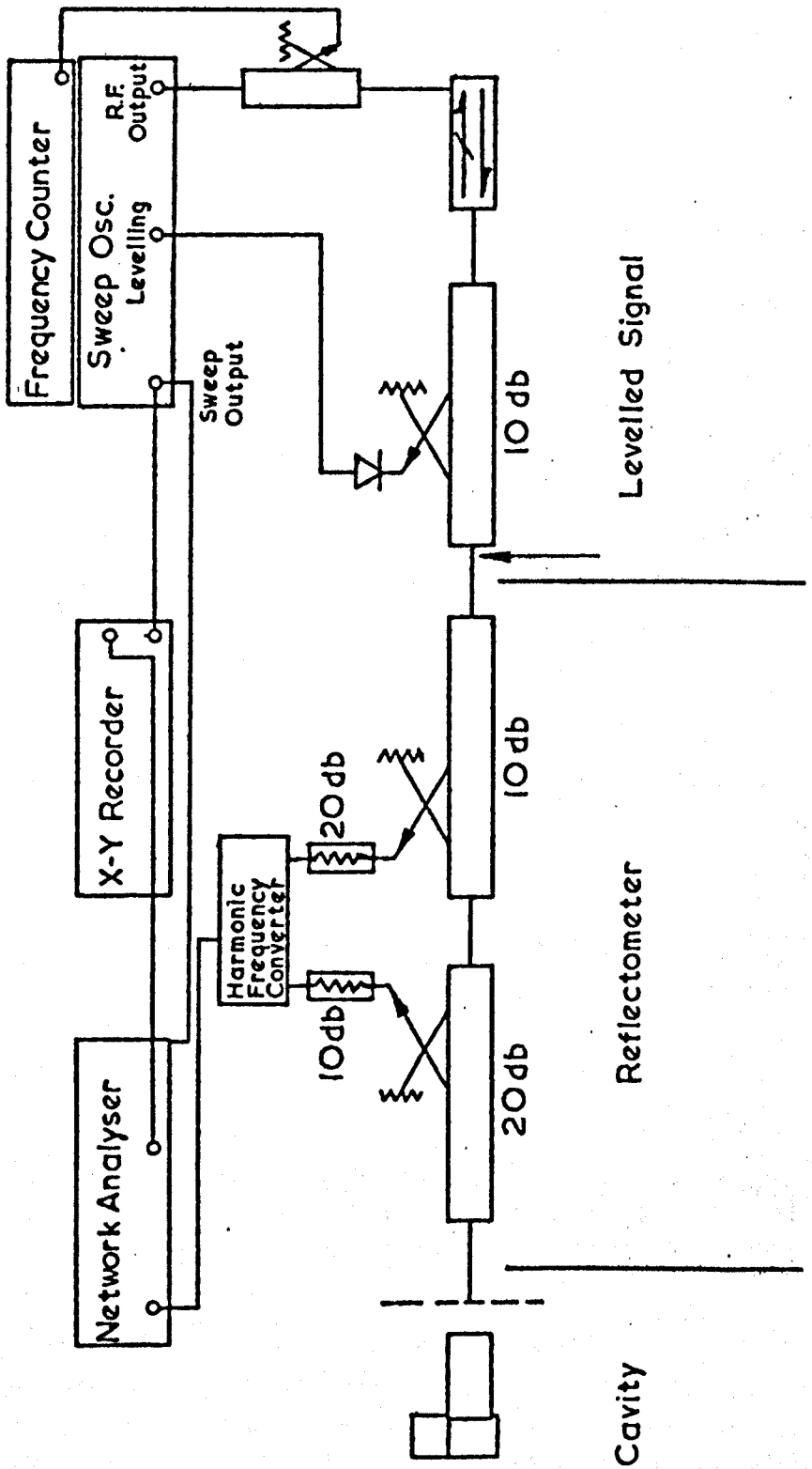


Figure 3. X-Band Reflectometer Unit

APPENDIX 18.

A microwave relaxation phenomenon in electron-beam-evaporated films of SiO.

J. Phys. D: Appl. Phys., 5, pp 1349-1351, (1972).

J. Phys. D: Appl. Phys., Vol. 5, 1972. Printed in Great Britain.

A microwave relaxation phenomenon in electron-beam-evaporated films of SiO

K MEHMET and M K McPHUN
Department of Engineering, University of Warwick, Coventry CV4 7AL

MS received 13 March 1972

Abstract. Measurements of the dielectric properties of electron-beam-evaporated films of silicon monoxide are presented. Results for frequencies from 1 MHz to 10 GHz show these films to have a dielectric relaxation phenomenon around 3 GHz. It is concluded that the relaxation is that due to torsional vibration predicted by Hirose and Wada.

1. Introduction

Vacuum-deposited silicon monoxide (SiO) films have found many uses in low-frequency electronic circuits. Therefore most of the investigations reported on SiO films were carried out at low frequencies. The few results published so far on SiO films at higher frequencies (Sobol and Hughes 1967, Park 1970) are insufficient to allow one to draw conclusions about the behaviour of this dielectric.

The work reported in this paper describes the high-frequency dependence of the dielectric properties of SiO films over four decades.

2. Experimental work

2.1. Preparation of thin films

The SiO films were deposited using an electron-beam evaporator. The bulk material was held in a carbon crucible with a source to substrate distance of 10 cm. The deposition conditions were as follows:

(i) The predeposition vacuum chamber pressure was 1×10^{-6} Torr.

(ii) The substrates were silver-plated Invar or Corning 7059 glass previously coated with silver, and were held at room temperature (20 °C).

(iii) A beam power of 150 W was used giving a deposition rate of 43 nm s^{-1}.

2.2. Measurement techniques

During the deposition the films were deposited simultaneously on several substrates. Subsequently one of the metal substrates with its film of SiO was used to make sandwich capacitors for the low-frequency measurements.

At frequencies up to 100 MHz the relative permittivity ϵ' and loss tangent, $\tan \delta$,

were obtained from measurements made on the sandwich capacitors. Commercially available impedance bridges were used.

The microwave properties of the films were measured using two separate techniques. One of these was a capacitively loaded coaxial cavity system (Mehmet and McPhun 1970) that can measure ϵ' and $\tan\delta$ at two frequencies between 1 and 10 GHz. The other technique used a symmetrical capacitively loaded microstrip resonator (Mehmet 1970); this system could measure at up to three different frequencies between 1 and 10 GHz.

3. Results and conclusions

The dielectric properties of vacuum-deposited films depend on various deposition parameters. To eliminate such ambiguities in the film properties, the SiO films were all deposited on silver-coated substrates, and during the same pump-down.

The measured results for one deposition are shown in figure 1. These results are consistent with those from other depositions. Here we see a dielectric relaxation phenomenon around 3 GHz. This is, we believe, the first time that such a relaxation frequency has been measured. The curve of $\tan\delta$ between 1·7 GHz and 4·5 GHz has been left broken, as it is quite possible that much higher losses exist in this region. Indeed the Debye equation predicts a maximum $\tan\delta$ ten times greater than that shown.

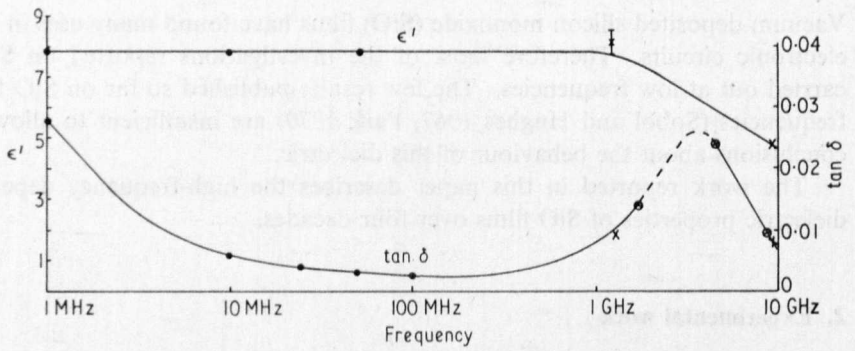


Figure 1. Dielectric properties of an SiO film. Measurement methods: ● RF bridges; × coaxial resonator; ⊗ microstrip resonator. (I represents measurement error.)

This relaxation was predicted by Hirose and Wada (1964) but above their measurement frequencies, and they attributed this relaxation to the 'local torsional vibration of Si—O chains'. Argal and Jonscher (1968) also predicted a loss peak to lie in the microwave frequency spectrum due to a two-centre electron-hopping mechanism. However, the large change in ϵ' observed can only be a lattice and not a carrier effect. Therefore our measured relaxation in electron-beam-evaporated SiO films is most likely due to local torsional vibration of the SiO chain as predicted by Hirose and Wada.

Acknowledgment

The authors would like to thank the Science Research Council (UK) for financial support.

References

Argal F and Jonscher A K 1968 *Thin Solid Films* **2** 185–210
Hirose H and Wada V 1964 *Jap. J. appl. Phys.* **3** 179–90
Mehmet K 1970 *PhD Thesis* University of Warwick
Mehmet K and McPhun M K 1970 *Proc. Conf. on Dielectric Materials, Measurement and Applications: IEE Conf. Publ. No* 67/70 pp 42–5
Park K T 1970 *PhD Thesis* University of Southampton
Sobol H and Hughes J J 1967 *IEEE Trans.* **MTT-15** 377–8

APPENDIX 19.

Thin film dielectric measurements.

An invited tutorial paper presented at the tutorial conference, Measurement of High Frequency Properties of Materials, N.P.L. Teddington, 27th - 29th March 1972. The complete text enclosed was subsequently printed in the book "High frequency dielectric measurement" edited by J. Chamberlain and G.W. Chantry, I.P.C. Science and Technology Press Ltd., 1973. pp 60-68.

Thin film dielectric measurements

M. K. McPhun and K. Mehmet*
Department of Engineering, University of Warwick, Coventry
Invited tutorial paper

1 INTRODUCTION

Microwave measurements on thin films of dielectrics are particularly difficult because the volume of the specimen is so minute. Also, these films are generally not self-supporting but are deposited on a substrate. In normal dielectric measurements, the process by which the dielectric specimen was made is not of very great concern to the person devising a measurement technique; this is not the case with thin films of dielectric, here the measurement technique and the process used for depositing the film are inherently bound together. It is therefore necessary to discuss briefly the methods used to deposit these films, and outline the influence the various methods have on the measuring techniques that may be used.

Table 1 shows very approximate temperatures experienced in the various common processes. In this Table the 'process temperature' is the temperature experienced by the dielectric material being deposited as the film, whereas the third column 'substrate temperature' is the temperature to which the substrate is subjected during some part of the deposition process. For our purposes the substrate may be either a dielectric or a metal and will very commonly be a thin metal layer already deposited onto a dielectric substrate. We shall consider these processes in turn:

Evaporation. The dielectric is heated, usually in a crucible, to a sufficiently high temperature to cause it to evaporate and be deposited with thermal velocities onto the substrate. This process is commonly used with SiO. For other materials its use is limited because the very high temperatures required are likely to cause dissociation of compounds, e.g. SiO_2 can lose oxygen.

RF sputtering. The dielectric forms a target which is bombarded with ions. Particles of dielectric are knocked off the target on a molecular scale, and are deposited at fairly high velocity onto the substrate. This is a much lower temperature process and cooling of both the target and the substrate are normally used. It has the advantage that compounds may be deposited without the risk of dissociation of the component parts.

Reactive sputtering. A metal target is sputtered in an oxidising atmosphere. This deposits the dielectric oxide onto the substrate. An example of this is tantalum which may be deposited as tantalum pentoxide. Again this is a fairly low temperature process.

Chemical. The dielectric is deposited from a gaseous compound onto a heated substrate. An essential part of this process is that the substrate should be heated to a sufficient temperature to cause the reaction to take place at its surface. A common example is the pyrolytic deposition of SiO_2 from silane (SiH_4). Note that in the Table the substrate temperature is shown as being much higher than the process temperature. This is because it is common with this process to densify the films deposited, by heating them to a much higher temperature after the deposition has been completed.

Electrolytic. A metal substrate must be used as an anode in an electrolyte, the dielectric film being formed on the anode as a result of electrolysis. Examples are the anodising of aluminium and the formation of a tantalum pentoxide layer on tantalum. This is essentially a low temperature process.

Table 1. Processes for depositing thin films of dielectric

Process	Process temperature (°C)	Substrate temperature (°C)
Evaporation	2000	200–800
RF sputtering	200	200
Reactive sputtering	200	200
Chemical	400	400–1200
Electrolytic	<100	<100

The normal range of film thicknesses lies between 0.2 μm and 20 μm though in this paper we will also discuss thicker films than these. It will be seen that accurate measurement of the thickness of the film is absolutely essential.

Often dielectric films will adhere well only to a limited number of metals, e.g. chromium and nickel; it is common practice therefore to use a *seeding layer* of one of these preferred metals as a glue placed between the substrate and the dielectric film. This seeding layer can have a large influence on the measuring technique used. With some processes it is not possible to deposit a stable dielectric film on a dielectric substrate without the prior deposition of a metal seeding layer. The thickness of the seeding layer would normally be only in the region of 0.02 μm.

* Dr Mehmet is now with Cossor Electronics Ltd., Harlow, Essex. The paper was given by Mr McPhun.

A typical application of a dielectric thin film is a microwave capacitor with a cross-section such as Fig.1. In this Figure the vertical scale is very much exaggerated but it is drawn approximately to scale.

In choosing a measurement method one must consider the constraints of the process and the specimens available: the temperatures involved, the chemical compatibility of the substrate with the process used, the size of specimens and the shape of substrates that are available. For instance, a measurement technique requiring a 100 mm diameter substrate could not be considered if the process used was not capable of depositing a uniform film over a diameter of 100 mm.

Secondly, one must consider why the measurement of $\hat{\epsilon}$ is required. The primary interest may be in the materials science of the dielectric films, or alternatively it may be in the application to which the films are being put. The material interest may present such questions as 'How do deposition conditions effect $\hat{\epsilon}$?', 'Does $\hat{\epsilon}$ change with film thickness?' or 'Is the dielectric effected by the seeding layer?'. This may then imply that a given process must be used but there may be a free choice of specimen shape or substrate material. On the other hand, for the application interest, we may simply be interested in making good capacitors and wish to separate the dielectric losses from those of the conductors. We may want to know whether the metal films of the capacitor are affecting the properties of the dielectric. This may then imply that we have a free choice of process, but the specimen type, for example, the substrate may be completely specified by the application.

An ideal measurement method will:

(a) Measure the dielectric film deposited on its normal substrate, the substrate having no effect on the measurement;
(b) Measure selected areas of film non-destructively;
(c) Measure with the E field in the same direction as will occur in the application.

It will be seen that we have a very long way to go to meet these ideals. The errors we discuss are one or two orders of magnitude higher than those presented in some other papers in these Proceedings, but in this field the accuracies that we can obtain are still a worthwhile achievement.

1.1 A note on perturbation theory

As a number of techniques to be discussed use the perturbation theory, some of the implications of this technique will be discussed here. It may be summarized as follows.

The resonant frequency and Q of an empty cavity are measured. The measurements are then repeated with a specimen in position in the cavity. The specimen properties are then given by

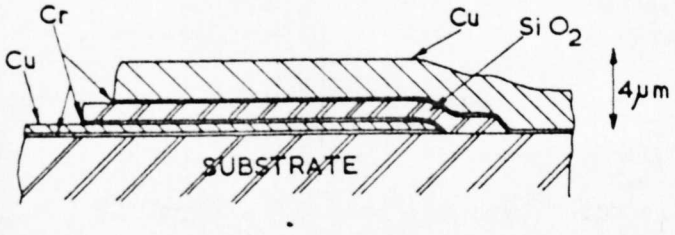


Fig.1. Cross-section of a microwave capacitor

$$\epsilon' = 1 - \frac{\Delta\nu}{\nu}\,\frac{V}{\Delta V}\,.\,F(x, y, z), \tag{1}$$

$$\epsilon'' = \frac{1}{2}\left(\frac{1}{Q_1} - \frac{1}{Q_0}\right)\frac{V}{\Delta V}\,.\,F(x, y, z), \tag{2}$$

where $\Delta\nu/\nu$ is the fractional shift in frequency upon insertion of the specimen, it is by convention negative and should ideally be less than 0.01; V is the volume of the cavity; ΔV is the volume of the specimen; Q_0 and Q_1 are the cavity Q's before and after the specimen is inserted; and F is a function of the cavity and specimen geometry, determined for instance by the proportion of electric field appearing within the specimen.

For the perturbation theory to apply $\Delta\nu/\nu$ must be small. For a given value of $\hat{\epsilon}$ a suitable value of $\Delta\nu/\nu$ may be obtained by varying the factors $V/\Delta V$ or F. Of course if $\Delta\nu/\nu$ is too small it will prove difficult to measure it accurately.

2 METHODS APPLICABLE TO SELF-SUPPORTED FILMS

Only the thicker films can be self-supporting and because of the greater volume of material present these films are generally the easiest to measure.

2.1 Park's method

This method is described fully in Section 3. It suffices here to note that it may also be used for self-supported films with thickness greater than 12 μm. For thicknesses greater than 50 μm or for the case where ϵ' is very high, a modification is required in order to keep the frequency shift within acceptable limits. For these cases either an annular specimen is used to increase the factor F, or the form of cavity shown in Fig.4 is used to increase $V/\Delta V$.

2.2 Mehmet's method [1]

This method utilizes a rectangular TE_{10n} cavity which is described fully in these Proceedings [2]. It has also been used to measure films of mica and glasses 30 to 100 μm thick. In these cases, the films extend over the whole area of the cavity cross section in order to minimise the factor F in equation (1) and hence maximise the frequency shift $\Delta\nu$.

2.3 Ahluwalia's method

Ahluwalia was interested in measuring self-supporting phospholipid membranes in the range 20 to 500 μm thick, and in order to test his method also measured other materials such as polyethylene. These films scarcely qualify to be called 'thin' but the technique used to minimise the factor F is of particular interest.

The two types of cavity used are shown in Fig.2. The first, a rectangular TE_{102} mode cavity, is very similar to that used by Mehmet [2]. The second, a circular TE_{012} mode cavity has the advantage of avoiding contact troubles at the sliding short, and gives a higher Q factor. All measurements were done at X-band. In both cases the film is placed tangentially to the E field, at a distance $\lambda_g/4$ from the short circuit plane where E is a maximum. In order to maximise the frequency

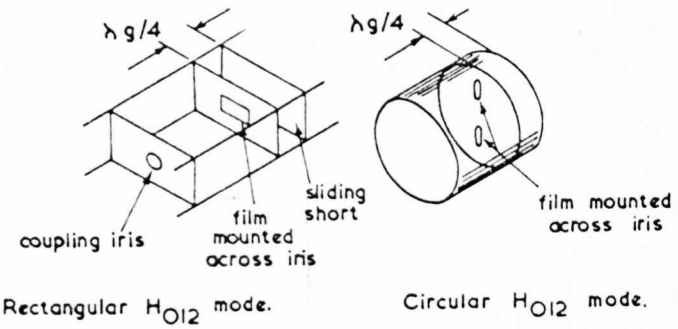


Fig.2. Ahluwalia's cavities

shift the electric field is concentrated in the plane of the specimen by mounting the film across a metal iris.

The film parameters are calculated using the perturbation theory. Without the iris, i.e. with the film extending over the whole cross-section, equation (1) reduces to

$$\epsilon' = 1 + \frac{\Delta\nu}{\nu_0}\,\frac{d}{t}, \tag{3}$$

where d = length of cavity and t = thickness of film.

With the film across the iris we have

$$\epsilon' = 1 + \frac{\Delta\nu}{\nu_0}\,\frac{d}{t}\,\frac{1}{\gamma}. \tag{4}$$

The factor γ can be determined experimentally by measurement of the same films with and without the use of the iris.

This is an accurate method, an error of ±1 per cent being obtained for the films with ϵ' in the region of 2 to 3. A minimum ϵ'' of 0.01 can be measured. There are, however, some difficulties.

Difficulties. The method is suitable only for thicker films that can be self-supporting, and mounting the film across the iris may be difficult. A sufficient area of film must be available to permit measurement without the iris, and this specimen must be sufficiently thick to enable an accurate value of γ to be obtained. E is tangential to the film and measurements with E perpendicular to the film are not possible.

3 METHOD APPLICABLE TO A FILM DEPOSITED ON AN INSULATING SUBSTRATE. PARK'S METHOD [4, 5]

A circular TE_{10n} cavity was used at X-band for films with a thickness t less than 10 μm. The value n = 1 was used as shown in Fig.3.

Two holes in the fixed end plate allowed coupling to rectangular waveguide. Not shown are the mode suppression devices which were used. The split cavity construction allows the specimen to be held without its thickness affecting the cavity length. This is achieved as follows. The specimen film is deposited onto an insulating substrate. This substrate is clamped between rings that can move longitudinally within an outer shell. The length of the cavity is determined by the movable short circuit end plate held in the outer shell. The specimen is tangential to the E field at a distance $\lambda_g^*/4$ from the short circuit. The

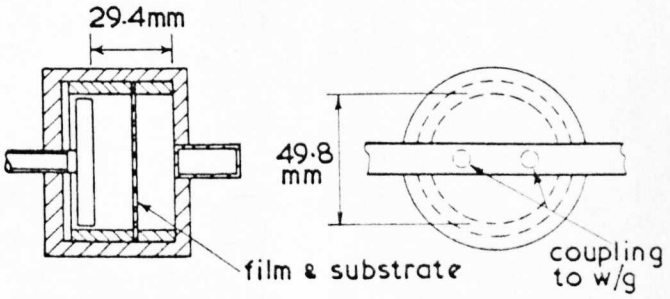


Fig.3. TE_{101} cavity for use at 8.96 GHz

perturbation theory is used to obtain the values of $\hat{\epsilon}$, the substrate being considered as part of the empty cavity. Thus the measurement technique consists of measuring the cavity centre frequency and Q, with the substrate in position. The substrate is then removed and the film deposited onto it. It is then placed back in the cavity and the measurements repeated. The method is capable of measuring very thin films, e.g. a 0.5 μm film of SiO deposited on a 12.5 μm mica substrate. For 3 μm films of polyethylene terephthalate (Melinex or Mylar) the errors quoted are $\Delta\epsilon'$ = ±3.4 per cent and $\Delta\epsilon''$ = ±17.5 per cent. A microwave bridge method of measurement was used and these figures were obtained without the use of an ultra-stable oscillator or a frequency counter. It is expected that given more elaborate measuring equipment the errors could be reduced considerably.

For thicker films, very lossy films or films with a high ϵ', the resulting frequency shift must be reduced. This may be achieved by measuring such films in a cavity with a larger value of n, such as the TE_{013} cavity shown in Fig.4. Thus in equation (1) $V/\Delta V$ is greater and $\nu/\Delta\nu$ is smaller for a given value of ϵ'.

A further measure employed for very lossy films is to photoetch the film into the form of an annulus to increase the factor F.

Advantages. The method can measure multi-layer specimens and promises high accuracy. Evaluation of the film parameters is nearly independent of the substrate thickness, special measures may be employed to give accurate results with lossy specimens or those having very large values of ϵ'.

Difficulties. The film must be deposited onto special circular substrates and must be uniform over a large area. The substrate must be of dielectric. It is not possible to use a seeding layer because the electric field is in the plane of

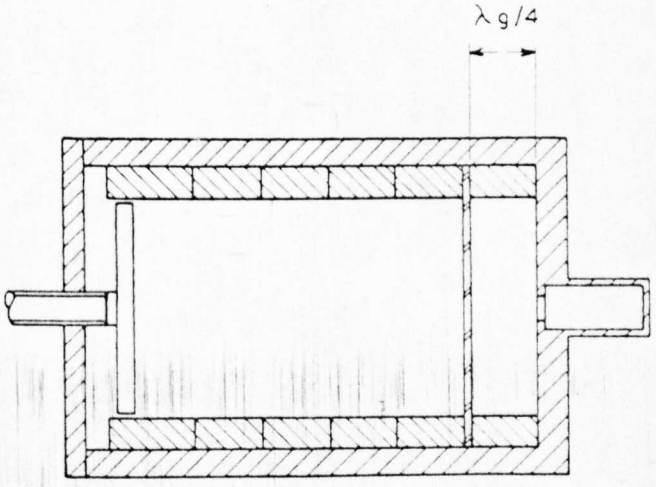


Fig.4. Park's cavity for thicker films or high ϵ'

the film, and conduction currents in the seeding layer would completely upset the measurements.

4 METHODS REQUIRING A FILM TO BE DEPOSITED DIRECTLY INTO A CAVITY

In the two methods to be described here, the empty cavity is first measured. Part of it is then removed and the film is deposited directly into this part of the cavity. The cavity is then reassembled and measured again. These two techniques have the following advantages and difficulties in common.

Advantages. The film may be measured with the E field perpendicular to the film. If required, the normal seeding layer may be used. A small area of film may be measured even at the lower microwave frequencies.

Difficulties. An existing film on an insulating substrate cannot be measured. Much more seriously, the cavity, or a part of it, is subjected to the film processing and must therefore be compatible with that processing, the temperatures used, etc. This also implies that a long time must elapse between the cavity being measured empty and being remeasured with the film present, and thus a very stable cavity is required.

4.1 Sobel's method

The cavity shown in Fig.5 uses a circular re-entrant TM_{011} mode at 2.9 GHz.

The factor F is minimised by depositing the film onto a small gap between the movable centre post and the end plate. A film thickness greater than 10 μm is required.

The calculation of film parameters utilizes an equivalent circuit approach. The cavity is considered as a radial transmission line giving an inductance which resonates with the central capacitor. The resonant frequencies with and without the film are made equal by adjusting the gap from d_1 without the film to $(d_2 + t)$ with the film. Precision spacers outside the cavity are used to set d_1 and d_2. For the case where $t \ll d_1 \ll r_1$, ϵ' is simply given by

$$\epsilon' = \frac{t}{d_1 - d_2}. \qquad (5)$$

Curves are plotted of resonant frequency *versus* d_1 and d_2 with and without the film as shown in Fig.6. Each point on the curve is the mean of many measurements.

From Fig.6 the quantity x is found, then

$$(d_1 - d_2) = t - x \qquad (6)$$

gives the value of $(d_1 - d_2)$ for use in equation (5).

Figure 7 shows the situation approximately to scale in the vertical direction, the value of x = 0.007 mm must be subtracted from t = 0.010 mm to give $(d_1 - d_2)$ = 0.003 mm. It is seen that a difference between similar values is involved and very accurate measurements of length and film thickness are absolutely vital. Nevertheless the method can give accuracies better than 10 per cent for a 1 μm film by taking the mean of many measurements. ϵ'' is found from the decrease in the Q of the cavity when the film is deposited in it. An equivalent circuit approach to the calculation was used similar to that described in section 4.2. In practice two cavities are used. The cavity for the measurement of ϵ' has the film deposited onto the end plate which is removable. For measurement of ϵ'' a fixed end plate is used to give a higher value of Q for the empty cavity, and the film is deposited onto the centre-post.

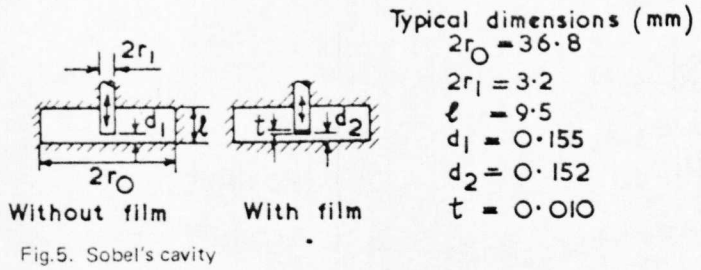


Fig.5. Sobel's cavity

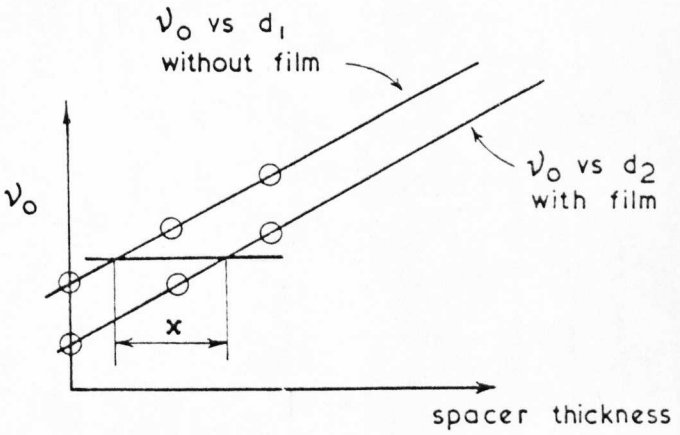


Fig.6. Measurements taken on Sobel's cavity

Advantages. These have been dealt with above.

Difficulties. In addition to those already outlined the method is extremely laborious, entailing the averaging of many measurements, and requires small differences to be used in the calculation. Because of the size of the cavity and the small movement required, the method is not thought to be practicable at X-band.

4.2 Mehmet's method [1, 6]

The cavity in this case is a TEM coaxial resonator, capacitively loaded at the centre. The film is placed in a small gap in the centre conductor where E has its greatest value. This type of cavity has been used for S-band and X-band measurements of SiO_2, SiO and Al_2O_3 with film thicknesses in the range 1 to 3 μm. Figure 8 shows the cavity drawn to scale, apart from the gap which is exaggerated.

Also shown in Fig.8 is the distribution of radial electric field along the cavity. This mode of resonance, with an extra $\lambda/2$ on each side of the gap, was used for the X-band measurements. The same cavity resonates in S-band in a mode without the extra $\lambda/2$ on each side of the gap.

The cavity material is Invar, silver plated and gold flashed. The cavity consists of two parts, a top and a base. The base is more complicated than the top as it contains a dumb-bell coupling to waveguide (for the S-band measurements a loop coupling to coax was used). Four tops were made to fit onto one base so that four specimens of film can be obtained from one deposition. The film is deposited directly onto the centre post of the simpler top part of the cavity, deposition being by means of evaporation or RF sputtering. Self-supporting 3 mm diameter discs with a thickness of 13 μm have also been measured.

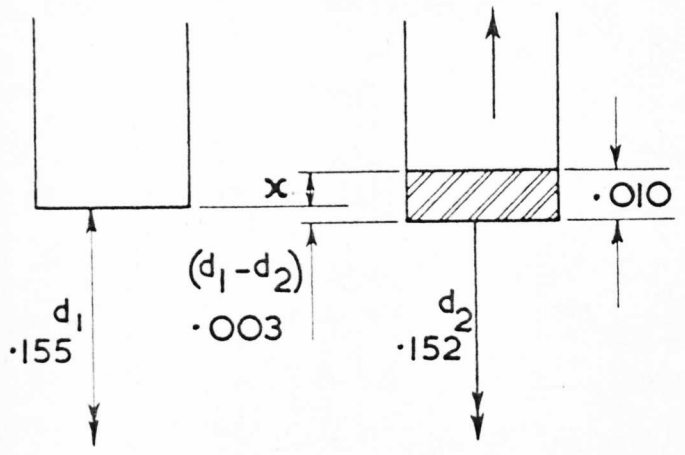


Fig.7. Gap dimensions

For measurement of ϵ', the optimum value of d_0/d_1 is approximately 6. However, for measurement of ϵ'' maximum Q is desirable, giving in this case an optimum value of $d_0/d_1 = 3.5$. The errors quoted later are for the cavity illustrated, with a compromise value of $d_0/d_1 = 5$.

The perturbation theory can be and has been applied, but a simpler measurement technique and more accurate values of ϵ'' result from using an equivalent circuit approach. The equivalent circuit of the empty gap is the π-section of Fig.9a. The values of the component parts of this π-section may be evaluated by carrying out measurements on the resonant frequencies of a number of different TEM modes in the cavity.

C_p and m can also be calculated from field theory. For the very small gaps used, C_p will normally be negligible giving the simplified equivalent circuit of Fig.9b, and for many purposes m may be neglected compared with unity also. With the film in place the equivalent circuit becomes as Fig.9c. From these equivalent circuits and straightforward transmission line theory we obtain

$$\epsilon' = x\left\{\left[(1+m)\frac{\nu_1 \tan k\nu_1}{\nu_2 \tan k\nu_2} - m\right]^{-1} + x - 1\right\}^{-1}, \quad (7)$$

where $k = 2\pi L/c$.

ϵ'' is calculated from the VSWR's at resonance, with and without the film [1]. This new technique* is easier and more accurate than that of measuring the cavity Q with and without the film. The principle is simply explained with reference to Figs 10 and 11.

Figure 10a shows the equivalent circuit of the empty cavity; with the film present $C_0(1+m)$ is replaced by the circuit of Fig.9c giving the series equivalent circuit of Fig.10b. Figure 11 shows the equivalent circuit of the cavity with its coupling represented by a mutual inductance, and the self inductance of the coupling absorbed into the cavity equivalent circuit and the transmission line.

Looking into the cavity without the film the impedance seen is

$$Z_0 = \frac{(\omega_0 M)^2}{r_1 r_2}; \quad (8)$$

with the film present the impedance seen is

$$Z_1 = \frac{(\omega_1 M)^2}{r_1' + r_2' + r_x'}, \quad (9)$$

where the primed quantities apply with the film present. Thus we have

$$\frac{Z_0}{Z_1} = \frac{S_1}{S_0} = \left(\frac{\nu_0}{\nu_1}\right)^2\left[\left(\frac{\nu_1}{\nu_0}\right)^{1/2} + \frac{r_x'}{r_2 + r_1}\right], \quad (10)$$

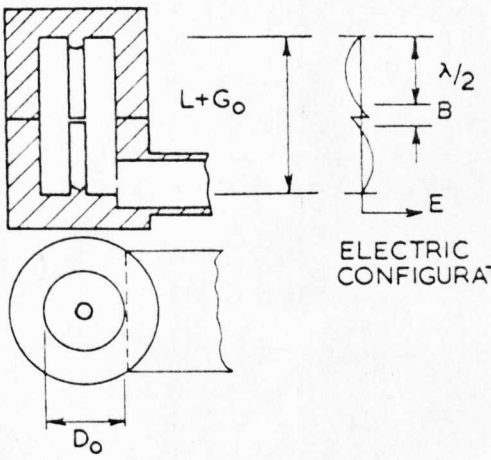
where S_0 is the input VSWR at resonance for the empty cavity and S_1 is the input VSWR of the cavity at resonance with the film in position.

Hence it may be seen that

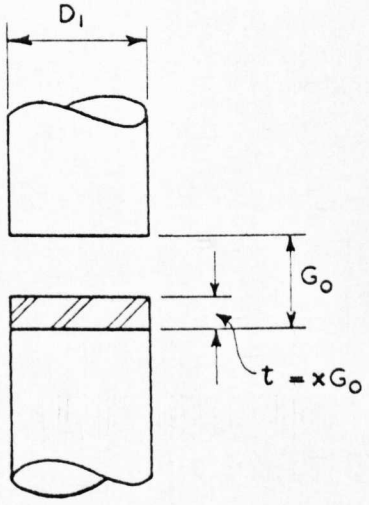
$$\epsilon'' \simeq \left(\frac{S_2}{S_1} - 1\right)\frac{(\epsilon')^2}{xQ_0K}, \quad (11)$$

where

$$K = 1 - \frac{2\pi}{k\nu_1 + \sin k\nu_1}.$$



a) CAVITY



b) GAP DIMENSIONS

Typical dimensions for

ν_0 = 9440 MHz
t = 1 to 5 μm
L = 32 mm
D_0 = 15 mm
D_0/D_1 = 5
G_0 = 30 μm

Fig.8. Mehmet's cavity

* Reference [14] discusses a similar measurement technique; it became available too late for inclusion in this discussion.

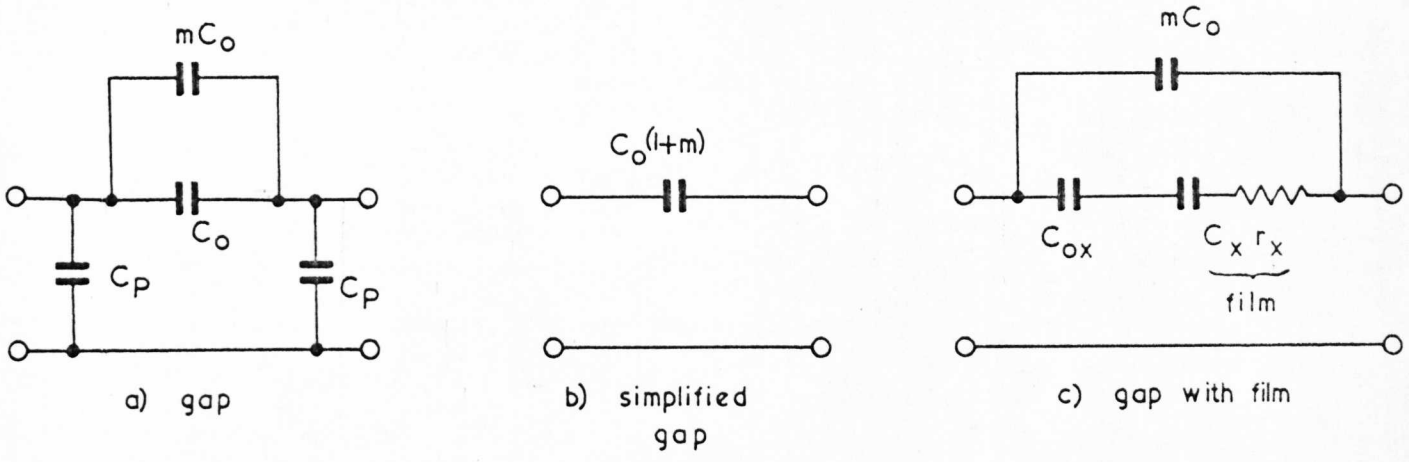


Fig.9. Equivalent circuits of gap

This equation is calculated with the assumption of uniform distribution of surface resistance within the cavity, i.e. no extra losses occurring at the split in the outer. The measured value of Q_0 is 93 per cent of the theoretical value.

The results obtained using this equivalent circuit approach were verified in two ways: firstly they were compared with the results calculated using the perturbation theory, and secondly they were verified by measuring known materials in the form of self-supported films. The materials used for construction of the cavity were chosen to be compatible with the dielectric film and with the processing required during its deposition, i.e. prolonged heating to 200°C in a vacuum. Measurements were carried out using a precision waveguide reflectometer [1, 6]. Q measurements were made to an accuracy of ±0.5 per cent. The measurement of resonant frequency was accurate to only ±0.1 MHz when the cavity was repeatedly dismantled and reassembled. The

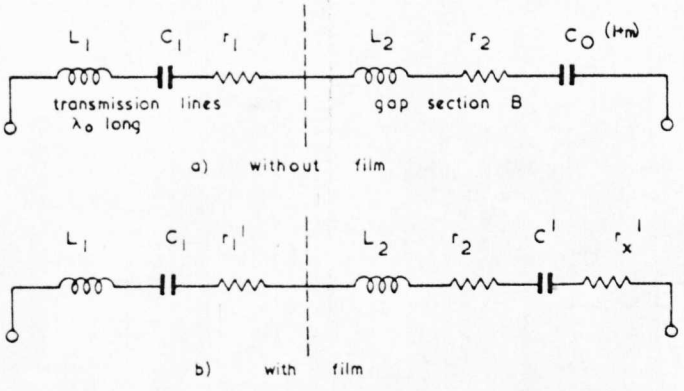


Fig.10. Equivalent circuits of cavity

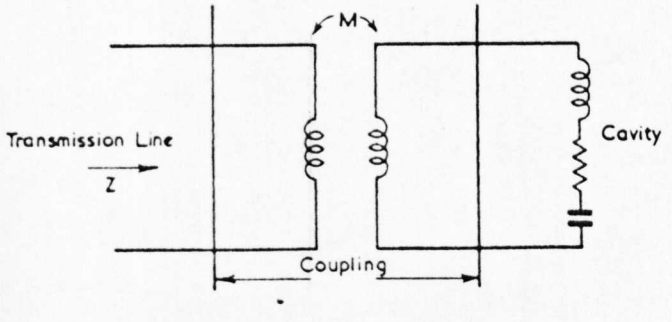


Fig.11. Equivalent circuit of coupling

Table 2. Errors obtained from a single measurement

Parameter	Equivalent circuit method	Perturbation method
ϵ'	±7.3%	±6.9%
ϵ''	±12%	±29%
tan δ	±9.4%	±29%

instability of the microwave oscillator used was significant and a more stable oscillator would give more accurate results. The ratios of VSWR's could be measured to an accuracy of ±0.16 per cent. The resulting errors in the material properties measured are shown in Table 2.

It is seen that a great improvement in accuracy is obtainable for the loss measurements by using the equivalent circuit approach rather than the perturbation theory. This is because the latter requires accurate measurements of small changes in narrow bandwidths.

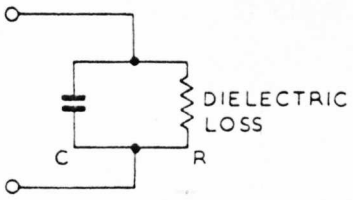
Advantages. Apart from those already outlined the method can give several measurement frequencies with one cavity by using several modes of resonance. Measurement of ϵ'' is simple and can give more accurate results than other techniques. An error in film thickness gives only a similar error in $\hat{\epsilon}$. The accuracy figures quoted could be improved with the use of a more stable oscillator.

Difficulties. The re-entrant shape of cavity is difficult to make to the close tolerances required.

5 INDIRECT METHODS

These methods involve using the film as the dielectric in a microwave sandwich capacitor as in Fig.1. One is then faced with the problem of modelling the capacitor and measuring its properties accurately.

Straightforward measurement of the capacitor impedance is not accurate enough due to the discontinuities involved in making the transition from the measuring system to the minute capacitor. We hope to overcome this using on-line computer correction techniques [8]. The severity of the problem of measuring the capacitor loss directly is shown in Fig.12. Microwave measurement of impedance will normally be made *via* a measurement of reflection coefficient.



$$Q_D = \frac{1}{\tan\delta} = \omega CR$$

ν = 7·96 GHz

tan δ = 0·01

FOR	C = 10 pF	C = 1pF	C = 0·1pF
$\frac{1}{\omega C}$ (Ω)	2	20	200
R (kΩ)	0·2	2	20
y	0·25 + j 25	0·025 + j 2·5	0·0025 + j0·25

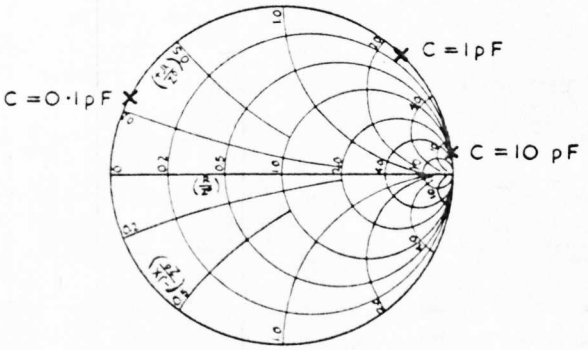


Fig.12. Admittance of lossy capacitor

In Fig.12 it is assumed that three capacitors with values of 0.1 pF, 1 pF and 10 pF are made using a dielectric with $\tan\delta = 0.01$. The admittances of these capacitors are shown tabulated at the frequency of 7.96 GHz and plotted on a Smith chart. It is seen that all three capacitors have an admittance lying practically on the rim of the chart, i.e. they are almost indistinguishable from lossless capacitors. Thus it is that present techniques involve using the capacitor to perturb a resonant transmission line; we are in fact back to a form of cavity measurement technique.

A satisfactory equivalent circuit model for the capacitor of Fig.1 must be used. In particular the conductor losses must be separated from those of the dielectric. Figure 13 shows a model for a capacitor with identical conducting films on both sides of the dielectric [9].

The measurements lead to a value of Q_x defined by

$$\frac{1}{Q_x} = \frac{1}{Q_c} + \frac{1}{Q_D}. \tag{12}$$

In this equation

$$Q_D = [\tan\delta]^{-1} \tag{13}$$

and the other relevant quantities are illustrated in Figs 12 and 13.

Calculation of ϵ'' depends on having a known value for R_S, the surface resistivity of the conducting films used. Correct use of skin depth in relation to film thickness is essential. Current work is aimed at measuring, with a separate cavity [10], R_S for the particular conducting films already used.

A microstrip transmission line is used to construct the resonator. In each case the components of the equivalent circuit are found from the changes in the resonant frequency and the Q, when the capacitor is added to the empty resonator. The main difficulties lie in the technique used for connecting the capacitor to the resonator. The capacitor is so minute that direct connection of a separately made capacitor into a resonator is almost impossible. Direct deposition of the capacitor into the resonator results in complex processing techniques being required if the empty resonator is to be measured before construction of the capacitor. Figure 14 shows to scale the plan view of such a 1 pF capacitor deposited across a gap in a microstrip line.

Various configurations of resonators may be used.

Figure 15 shows plan views of the top conductor of the microstrip line for three methods. In Fig.15a, Mehmet's resonator, the symmetrical microstrip line is terminated at both ends in a short circuit. Coupling is achieved at one end by feeding a current through the short circuit [13]. The resonance of reflection is measured. Michie's resonator [11], shown in Fig.15b, again requires measurement of resonance of reflection. In this case the line is terminated in a short circuit at one end but the input power is obtained *via* a capacitative coupling to the electric field This has the advantage that the coupling can be closely controlled, so that an optimum value for accurate measurements may be chosen. Hughes' resonator [12], shown in Fig.15c, utilizes a transmission method of measuring resonance. A capacitative coupling to the electric field is used at both input and output. An essential part of this method is that the coupling should be sufficiently weak not to affect the measurements.

All three methods use an equivalent circuit approach to the calculation of the material properties, and not the perturbation technique. Thus very large changes in frequency and Q factor can be permitted, and this should permit accurate measurements to be obtained. However, the accuracy of the values obtained for the properties of the dielectric material can never be greater than the accuracy of the equivalent circuit model used.

Advantages. The method has the potential of giving accurate measurements, particularly for ϵ''. It permits measurement of the dielectric properties *in situ*.

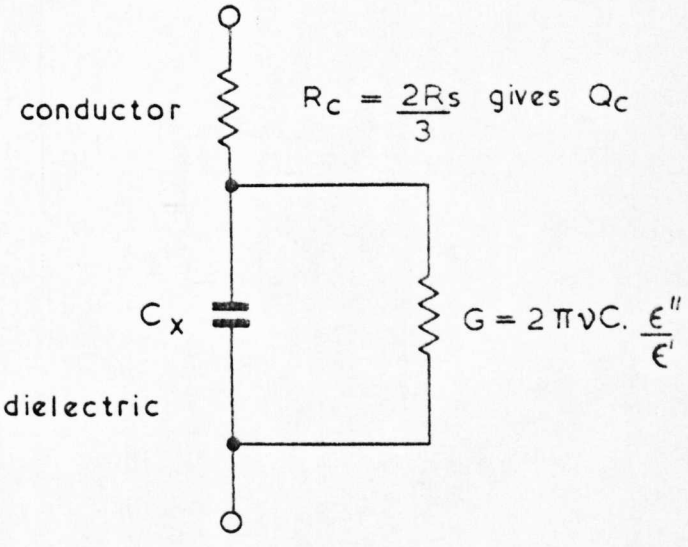


Fig.13. Model of capacitor

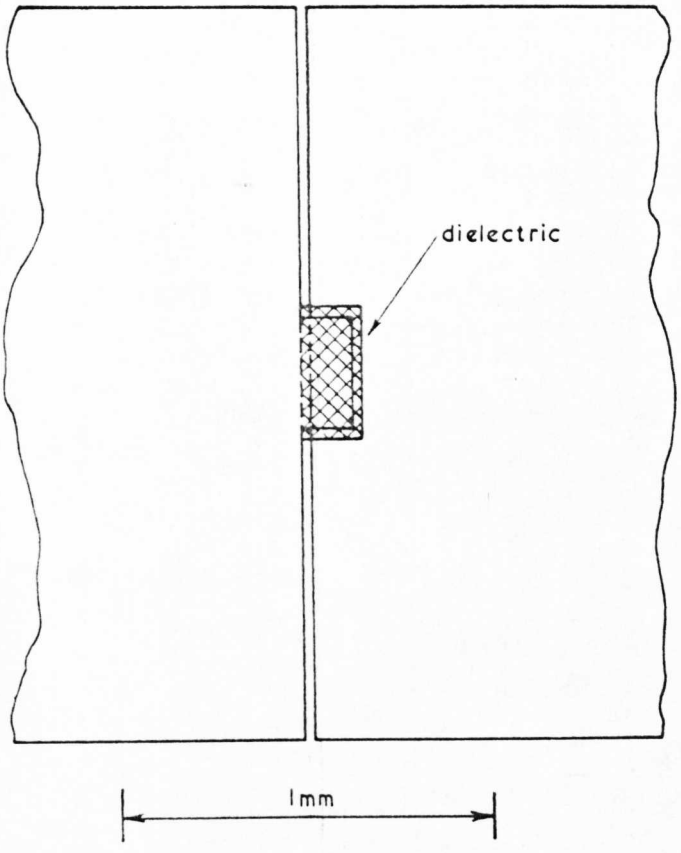


Fig.14. 1 pF capacitor deposited across gap in microstrip transmission line

Difficulties. The typical capacitor size is a few hundred μm square, giving rise to discontinuities when connected to the microstrip line. The resonator must withstand the film processing techniques if the capacitor is deposited directly onto the resonator.

6 CONCLUSIONS

In the space of this short review it has been necessary to gloss over many important factors and approximate freely for the sake of conciseness. The reader is strongly advised to consult the source references before contemplating measurements of his own.

Of the methods described it seems clear that Park's method is capable of giving the most accurate results providing the conditions necessitated can be satisfied, i.e. for X-band measurements a uniform film 50 mm diameter is required deposited on an insulating substrate without a seeding layer.

For small areas of film at microwave frequencies up to 12 GHz, with the electric field perpendicular to the film being measured, and with the use of a seeding layer permitted, then we recommend the use of Mehmet's method. We expect that the indirect method, i.e. measuring a capacitor made with the dielectric film, will yield more accurate results in the future, but work on these methods is still in progress.*

It is clear that much work remains to be done in this field.

* References [15] and [16] also appeared too late to be discussed. They are very relevant, and present the most recent results on this subject.

REFERENCES

1 Mehmet, K. Ph.D. Thesis, University of Warwick (1970).

2 Mehmet, K. and McPhun, M. K. this volume p.69.

3 Ahluwalia, H. P. S., Boerner, W. M. and Hamid, M. A. *Proc. I.R.E.E. (Australia)* Vol.32 (6), 253 (1971).

4 Park, K. T. Ph.D. Thesis, University of Southampton (1970).

5 Park, K. T. and Stachera, H. Conference on 'Dielectric Materials Measurements and Applications' *I.E.E. Conference Publication* No.67/70, 258 (1970).

6 Mehmet, K. and McPhun, M. K. Conference on 'Dielectric Materials Measurements and Applications' *I.E.E. Conference Publication* No.67/70, 42 (1970).

7 Sobel, H. and Hughes, J. J. *I.E.E.E. Trans.* MTT-15 (6), 377 (1967).

8 Shurmer, H. V. *Radio and Electronic Engineer* Vol.41, 357 (1971).

9 deBrecht, R. and Caulton, M. *GMTT International Microwave Symposium,* California (May 1970).

10 Butlin, R. S. and McPhun, M. K. *Electronics Letters* Vol.8, 637 (1972).

11 Michie, D. Ph.D. Thesis, University of Warwick (1973).

12 Hughes, J. J., Napoli, L. S. and Reichert, W. F. *Electronics Letters* Vol.5, 535 (1969).

13 Mehmet, K. and McPhun, M. K. *Electronics Letters* Vol.8, 165 (1972).

14 Eldumiati, I. I. and Haddad, G. I. *I.E.E.E. Trans.* MTT-20 (2), 126 (1972).

15 Rzepecka, M. A. and Hamid, M. A. K. *I.E.E.E. Trans.* MTT-20 (1), 30 (1972).

16 deBrecht, R. E. *I.E.E.E. Trans.* MTT-20 (1), 41 (1972).

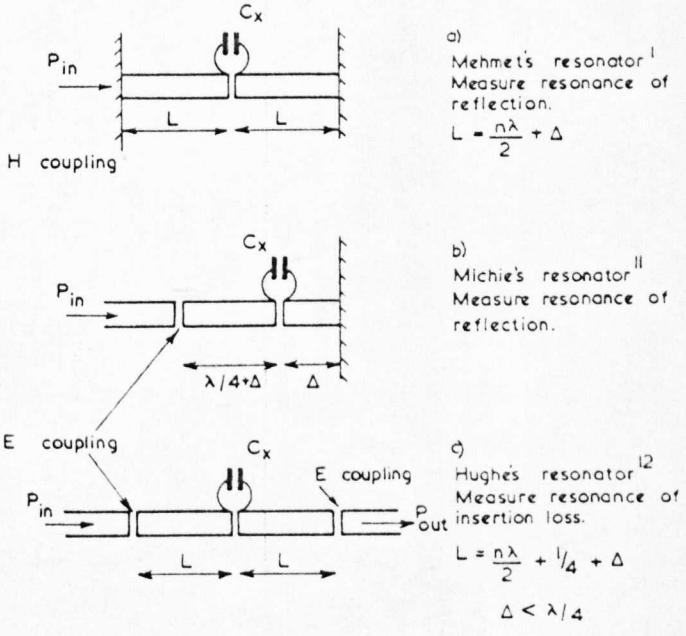


Fig.15. Microstrip resonators

DISCUSSION

DR M. KENT (Torry Research Station) asked what was the Q of the microstrip cavity and MR McPHUN replied that it was 700 but that this was quite adequate. DR G. J. HILL (ERA, Leatherhead) asked what were the lowest values of permittivity that could be measured at 2.8 GHz using Park's method. He also asked if it were possible to measure high permittivity, say of the order 300. MR McPHUN replied that Park's method was not really suitable for low permittivity measurements at 2.8 GHz due to the large area of film required at that frequency. However, for high permittivity materials, an annulus or small disc of the material would be deposited. Thus permittivities of 300 should be measurable with ease, even at 2.8 GHz.

PROF. J. LAMB (University of Glasgow) said that he had used the Mehmet cavity and found that it had two modes whose frequencies are very conveniently placed. One of these gives an electric field which is concentrated and radial at the gap. One could, therefore, use the dielectric substitution method.

DR R. J. COOK (NPL) asked if with high permittivity materials different results were obtained for different specimen positions in the cavity. MR McPHUN replied that so far he had made no attempt to test such effects since he had been working to reduce errors of known origin. In any case, he had so far confined his measurements to relatively low permittivity materials ($\epsilon' \sim 3$ to 10), for example alumina.

APPENDIX 20.

Measurement of the properties of rectangular plates using a rectangular cavity.

A paper presented at the tutorial conference, Measurement of High Frequency Properties of Materials, N.P.L. Teddington, 27th - 29th March 1972. The enclosed text was subsequently printed in the book "High frequency dielectric measurement" edited by J. Chamberlain and G.W. Chantry, I.P.C. Science and Technology Press Ltd., 1973, pp 69-72.

Measurement of the properties of rectangular plates using a rectangular cavity

K. Mehmet* and M. K. McPhun
Department of Engineering, University of Warwick, Coventry

1 INTRODUCTION

Various measurement techniques are employed for the measurement of the microwave dielectric properties of materials. Some of these techniques and their limitations are described by other authors in these Proceedings.

Our problem was to determine the variation of ϵ' and $\tan\delta$ across the area of alumina substrates having a thickness of about 0.6 mm.

The measurement approach presented uses the perturbation theory applied to a rectangular cavity. The theoretical approach, experimental techniques and the errors will be discussed in the following sections.

2 THEORETICAL APPROACH

The perturbation technique has been used in many investigations especially on cavities [1, 2]. The relationship between the parameters of an unperturbed cavity and the same cavity but perturbed by a nonmagnetic dielectric material is given by

$$\frac{\Delta\nu}{\nu_1} + j\Delta\left(\frac{1}{2Q}\right) = -\epsilon_0(\hat{\epsilon} - 1)\,\frac{\iiint_v \hat{E}_i\hat{E}_0\,dV}{4\mathscr{E}}, \qquad (1)$$

where

$\Delta\nu = \nu_2 - \nu_1$,

ν_1 = resonant frequency of the unperturbed cavity,

ν_2 = resonant frequency of the perturbed cavity.

$$\Delta\left(\frac{1}{2Q}\right) = \frac{1}{2}\left(\frac{1}{Q_2} - \frac{1}{Q_1}\right),$$

Q_1 = unloaded Q-factor of the unperturbed cavity,

Q_2 = unloaded Q-factor of the perturbed cavity,

$\hat{\epsilon}$ = complex relative permittivity of the dielectric specimen,

$\hat{E}_i$ = electric field inside the specimen,

$\hat{E}_0$ = electric field in the same plane but with the dielectric specimen removed,

V = volume of the specimen, $\mathscr{E}$ = energy stored in the cavity.

Now this principle will be applied to a rectangular cavity as in Figure 1a.

Assuming resonance in the TE_{10n} mode, the field distribution in the cavity is given by [3]

$$\left.\begin{aligned}
\hat{E}_y &= \hat{E}\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{n\pi z}{d}\right),\\
\hat{H}_x &= -j\frac{\hat{E}}{\eta}\,\frac{\lambda}{2d}\sin\left(\frac{\pi x}{a}\right)\cos\left(\frac{n\pi z}{d}\right),\\
\hat{H}_z &= j\frac{\hat{E}}{\eta}\,\frac{\lambda}{2a}\cos\left(\frac{\pi x}{d}\right)\sin\left(\frac{n\pi z}{d}\right),
\end{aligned}\right\} \qquad (2)$$

where $\eta = \nu\mu\lambda_g$

and λ_g = guide wavelength for the TE_{10} mode.

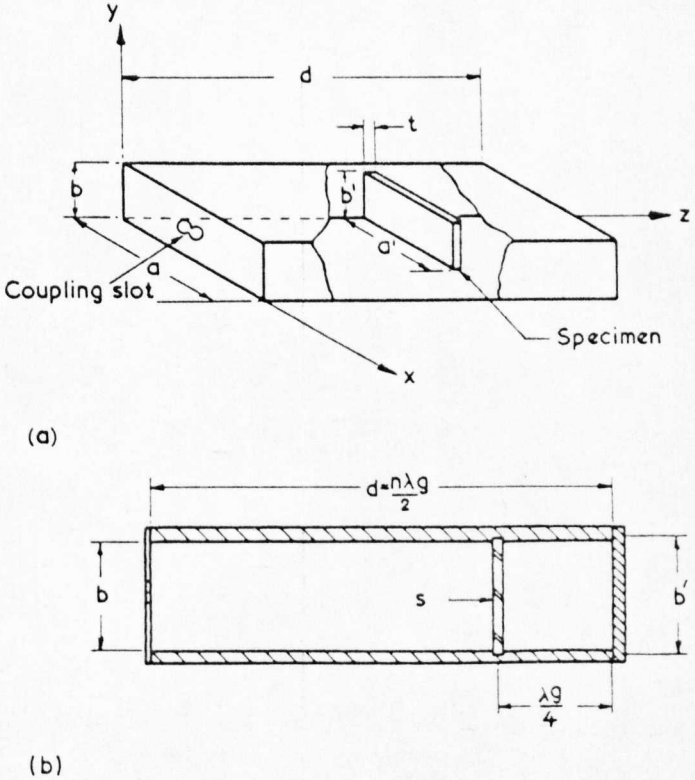


Fig.1. (a) Cavity loaded with dielectric specimen
(b) Cross section of cavity with specimen in recessed slot

* Dr Mehmet is now with Cossor Electronics Ltd., Harlow, Essex. The paper was given by Dr Mehmet.

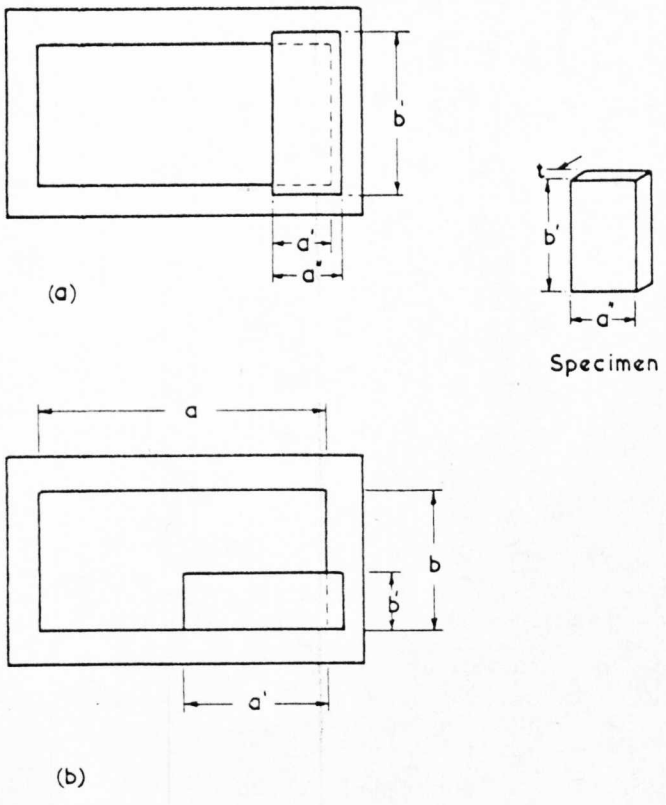


Fig.2. Cross-section of loaded cavity showing specimen dimensions

At a place $\lambda_g/4$ from the shortcircuit plane the field distributions are given by

$$\left.\begin{aligned} \hat{E}_y &= \hat{E}\sin\left(\frac{\pi x}{a}\right), \\ \hat{H}_x &= 0, \\ \hat{H}_z &= j\frac{\hat{E}}{\eta}\frac{\lambda_g}{2a}\cos\left(\frac{\pi x}{a}\right). \end{aligned}\right\} \tag{3}$$

A thin dielectric sheet is placed across the cavity at this point as shown in Figure 1a. Under these conditions

$$\begin{aligned} \hat{E}_i &= \hat{E}_0 \\ &= \hat{E}_y = \hat{E}\sin\left(\frac{\pi x}{a}\right). \end{aligned} \tag{4}$$

Substituting these values in equation (1) we obtain

$$\frac{\Delta\nu}{\nu_1} + j\Delta\left(\frac{1}{2Q}\right) = \frac{-\epsilon_0(\hat{\epsilon} - 1)\int_0^t\int_0^{b'}\int_0^{a'} \hat{E}^2\sin^2\left(\frac{\pi x}{a}\right)dx\,dy\,dz}{2\epsilon_0\int_0^d\int_0^b\int_0^a \hat{E}^2\sin^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{n\pi z}{d}\right)dx\,dy\,dz} \tag{5}$$

where a, b and d are the cavity dimensions and a', b' and t are the specimen dimensions as shown in Figs 1a and 2.

Integrating and substituting the respective values of x, y and z we obtain

$$\epsilon' \simeq 1 - P\left(\frac{\Delta\nu}{\nu}\right) \tag{6a}$$

and

$$\epsilon'' = P\Delta\left(\frac{1}{2Q}\right), \tag{6b}$$

where

$$P = \frac{abd}{a'b't}\left[1 - \frac{a}{2\pi a'}\sin\left(\frac{2\pi a'}{a}\right)\right]^{-1}$$

and $\Delta\nu$ is negative.

We also use the loss tangent $\tan\delta$ which, from equations (6a) and (6b), is given by

$$\tan\delta = \frac{\epsilon''}{\epsilon'} = \frac{P\Delta(1/2Q)}{\epsilon'}. \tag{7}$$

These equations are used to calculate the dielectric properties of the specimens.

3 EXPERIMENTAL WORK

The cavities were made from WG16 waveguides. Two of these cavities which resonated at about 9.5 GHz had the following characteristics:

Cavity 1: TE_{102} mode, Q_1 = 6000 (≈7000 when polished)
Cavity 2: TE_{104} mode, Q_1 = 6800 (≈7400 straight after silver plating)

These cavities were split at $\lambda_g/4$ from the shortcircuit plane as shown in Fig.1b and a recessed slot was prepared to accommodate the specimens which were placed in the cavity as shown in Fig.2a and 2b. Note that at this plane the wall currents are entirely transverse, i.e. parallel to the split. The cavity was excited through a dumbell slot and its parameters, namely Q and ν were measured using an X-band precision reflectometer system [4].

Figure 3 shows the dismantled cavity.

The experimental procedure was to measure the empty cavity parameters, then insert the specimen and measure

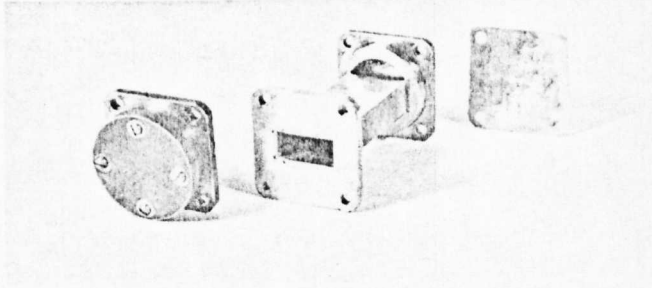


Fig.3. The dismantled cavity parts

Table 1

Material	Dielectric properties: Measured ϵ'	tan $\delta \times 10^3$	ν (GHz)	Reported by others ϵ'	tan $\delta \times 10^3$	ν (GHz)	Ref
ALUMINA							
Coors 99.5	10	0.13	9.5	9.8	0.1	10	
091	9.54	0.8					
092	9.57	1.3					
093	9.40	0.8					
094	9.55	1.4					
221	9.81	0.3					
223	9.86	0.58					
224	9.74	0.58					
101	9.36	0.8					
102	9.57	1.31					
103	9.55	1.22					
	9.68			9.6		0.001	[5]
	9.41			9.64		0.001	
Expanded Polystyrene	1.06	3					
Ruby Mica (70)	5.8	0.4		5.4	0.3	10	[6]
Amber Mica (60)	6.0	2.3		6.1	2.2	9.5	[4]
Soda Lime Glass	6.1	16.0		6.1	15.0	9.5	[4]

the new loaded cavity parameters. The dielectric properties of the specimen were then calculated using equations (6a), (6b) and (7). In this investigation various materials were used. The measured results and the values quoted by other workers are shown in Table 1.

4 MEASUREMENT ERRORS

The errors of the measurement system can be summarized as follows.

4.1 Systematic errors

These are due to calibration errors, system faults and biasing, and proper design can eliminate them. If they are not eliminated they will introduce an error which will give a bias equal in magnitude in all of the measured values. As an illustration consider a dielectric specimen mounted as in Fig.2b. Measured results for this arrangement are shown in Table 2. Here we have a systematic error. It arises from the discontinuity of the electric field owing to the air gap between the specimen and the top of the cavity. This error can be removed by adopting the approach shown in Fig.2a (and Fig.1b). Note that a' and not a'' should be used in (6b).

Table 2. Systematic errors

Material	Measured ϵ'	ν (GHz)	Measured by others ϵ'	ν (GHz)	Difference %
1	8.92	9.5	9.49	9.3	−5.6
2	9.17		9.6		−4.5
3	8.47		9.06	10.0	−6.5
4	9.28		9.8	10.0	−5.3

4.2 Random errors

These can be estimated using statistical methods [4]. The probable fractional error $\delta\epsilon'/\epsilon'$ in ϵ' as found in the present technique can be estimated from the equation

$$\frac{\delta\epsilon'}{\epsilon'} = \pm 0.9\left[\left(\frac{\delta b'}{b'}\right)^2 + \left(\frac{\delta t}{t}\right)^2 + 8\left(\frac{\delta a'}{a'}\right)^2 + 2\left(\frac{\delta \nu}{\nu}\right)^2\right]^{1/2}, \quad (8)$$

where $\delta b'/b'$, . . . etc. are the maximum fractional spread-values. The errors in our measurements are shown in Table 3, which gives $\delta\epsilon'/\epsilon' \simeq \pm 0.014$ (1.4%); the measured spread values were lower than this value.

Table 3. Random errors

Variables	Average values	Maximum spread
a (mm)	4.0	±0.01
b (mm)	10.8	±0.02
t (mm)	0.63	±0.003
ν_1 (MHz)	9467.3	±0.3
ν_2 (MHz)	9446.0	±0.3

5 CONCLUSIONS

The measurement technique presented here is quite simple and practical and the error analysis of the measuring system showed that errors within ±1.4 per cent in ϵ' can be achieved. It will be seen in section 4.2 that the most important error parameter is a'. This can be reduced by extending the specimen (if possible) across the area of the cavity. In order to satisfy the perturbation theory and keep $\Delta\nu/\nu$ smaller than 0.005, it is then necessary to reduce the ratio of the volume of the specimen to that of the cavity, i.e. to increase the length of the cavity. For this work however, it was desirable to work with small specimen areas.

ACKNOWLEDGEMENT

This work was carried out in the Department of Engineering, University of Warwick. The authors wish to thank the Royal Radar Establishment for the financial support and P. J. Smith for the construction of the cavities.

REFERENCES

1 Waldron, R. A. *Proc. I.E.E.* Vol.107C, 272 (1960).

2 Gos'kov, P. I. *Sov. Phys. J.* Vol.8 (4), (1965).

3 Ramo, S., Whinnery, J. R. and vanDuzer, T. *Fields and waves in communication electronics,* p.542 John Wiley, New York (1965).

4 Mehmet, K. Ph.D. Thesis, University of Warwick (1970).

5 Hill, G. J. private communication.

6 Montgomery, C. G., Dicke, R. H. and Purcell, E. M. *Principles of microwave circuits,* p.382 McGraw-Hill, New York (1948).

APPENDIX 21.

Surface-resistance measurements of thin conducting films at 10 GHz.

Electronics Letters 8, 26, pp 637-639, (28th Dec. 1972).

pp. 637–639

SURFACE-RESISTANCE MEASUREMENTS OF THIN CONDUCTING FILMS AT 10 GHz

Indexing terms: Calibration, Cavity resonators, Microwave measurements, Resistance measurements, Surface phenomena, Thin films

The letter describes a technique for calibration of an H_{011} cavity for surface-resistance measurements of thin-film and bulk samples at 10 GHz. After calibration and insertion of the sample, it is only necessary to measure the return loss or v.s.w.r. at resonance. The conductivity and surface resistance of the sample may then be read directly from a graph.

The performance of many microwave integrated circuits (m.i.c.s) is critically dependent on the microwave surface resistance of thin metal films, often built up by electroplating. Experience shows that this surface resistance cannot be predicted from direct-current or low-frequency measurements. This letter describes a technique for calibration of a cylindrical H_{011} cavity for surface-resistance measurements of thin-film and bulk samples at 10 GHz.

The thin film is measured after deposition onto its normal substrate. This forms one end plate of the H_{011} cylindrical cavity, as shown in Fig. 1. A dielectric substrate is supported only by its corners, so that it has no conducting backing. The other end plate forms an integral part of the cavity. This is ideally suited for surface-resistance measurements, as for the pure mode there are no current paths between the bore of the cavity and the end plates. Thus contact resistance of a detachable end plate will have no effect on the Q factor of the cavity. However, associated with the H_{011} mode is the the degenerate E_{111} mode. Interference from this mode may be eliminated by putting annular grooves in the solid end of the cavity, as shown. Provided that there is negligible field penetration into the grooves, the effect of the grooves on the H_{011} mode may be obtained by considering the grooved end of the cavity to have an effective conductivity σ_2. The Q factor of the cavity is now given by

$$\frac{1}{Q} = \frac{1}{Q \text{ of bore}} + \frac{1}{Q \text{ of grooved end}} + \frac{1}{Q \text{ of sample end}}$$

i.e.

$$\frac{1}{Q_S} = \frac{K\lambda^2}{\pi\lambda_g \delta_1 \sigma_1} + \frac{2\lambda^2}{\pi\lambda_g^2 Z_w \delta_2 \sigma_2} + \frac{2\lambda^2}{\pi\lambda_g^2 Z_w \delta_s \sigma_s} \quad (1)$$

where λ is the free-space wavelength at resonant frequency $\omega_0/2\pi$, λ_g is the guide wavelength at resonant frequency $\omega_0/2\pi$, Z_w is the wave impedance of H_{01} circular waveguide mode, σ_1 is the conductivity of bore material, δ_1 is the skin depth in bore material, σ_s is the conductivity of sample, δ_s is the skin depth in sample, σ_2 is the effective conductivity of grooved end, δ_2 is the effective skin depth in grooved end, Q_s is the Q factor of cavity with sample, K is the $\alpha\delta_1\sigma_1$ and α is the attenuation constant of H_{01} circular waveguide mode.

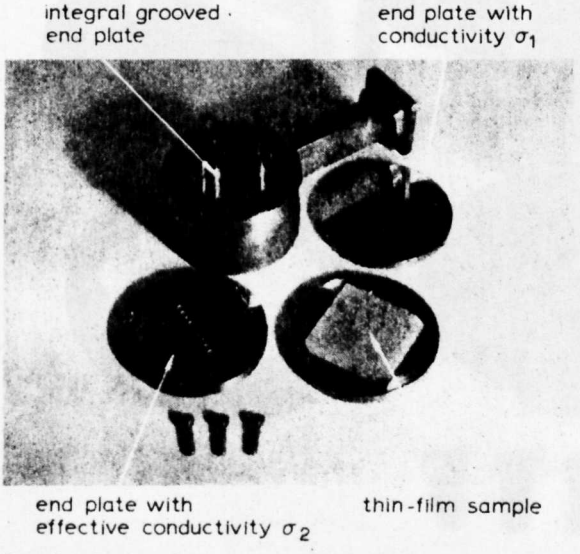


Fig. 1 *Test cavity showing calibration end plates and thin-film sample*

Before the surface resistance of the sample can be found, the surface resistance ($1/\delta\sigma$) of the bore and grooved end plate must be known. This may be found by measuring the cavity Q factor, first with a detachable end plate of conductivity σ_1 and then with an end plate having identical grooves to the opposite end and an effective conductivity σ_2. Thus, three Q factor measurements are required to obtain the surface resistance of the sample. However, these measurements may be greatly simplified, as described below.

Consider the cavity to be represented by a series-tuned circuit with resistance R, coupled via a mutual inductance M[1]. For the cavity, $Q = \omega_0 L/R$, and ω_0 and L may be assumed to remain constant for all conducting samples. Any two Q factor measurements may then be related by

$$\frac{Q_1}{Q_2} = \frac{R_2}{R_1} = \frac{S_2}{S_1}$$

where Q_1 and Q_2 are the cavity Q factors for resistances R_1 and R_2, respectively, giving v.s.w.r.s at resonance of S_1 and S_2.

Therefore, Q_1 having been measured, S_1 and S_2, Q_2 may be found. In practice, it has been found more convenient to measure the return loss.

On returning to eqn. 1, we find that Q_S may be written as a function of Q_1 and S_1, the cavity Q factor and the v.s.w.r. for a detachable end plate of conductivity σ_1, and S_S, the v.s.w.r. with the sample end plate. Rearranging the equation gives

$$\frac{1}{\delta_S \sigma_S} = \frac{\pi\lambda_g^2 Zw}{2\lambda^2 Q_1(S_1/S_S)} - \frac{K\lambda_g Zw}{2\delta_1 \sigma_1} - \frac{1}{\delta_2 \sigma_2}$$

If Q_1, S_1, $\delta_1\sigma_1$ and $\delta_2\sigma_2$ are known, a calibration curve of $1/\delta_s\sigma_s$ against S_S may be drawn. Return-loss measurements have been made on the present cavity to obtain the calibration curve of Fig. 2A, which shows the variation of surface resistance and conductivity of the sample with return loss. Careful analysis of errors yields a maximum error in R_s of $\pm 0{\cdot}0012\,\Omega$. Thus, the cavity having been calibrated, all subsequent surface-resistance measurements may be made by measurement of the return loss at resonance alone.

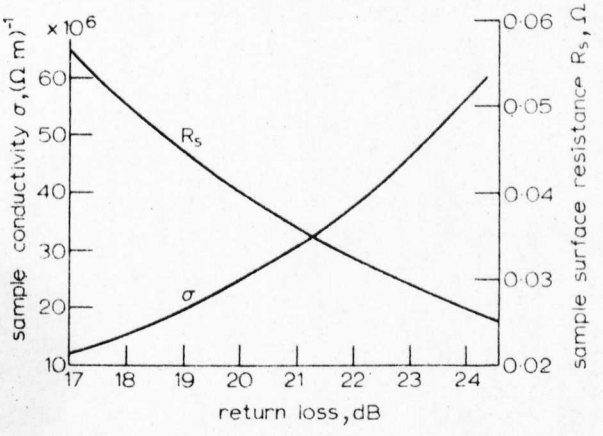


Fig. 2A *Calibration curve showing variation of sample conductivity and surface resistance with return loss*

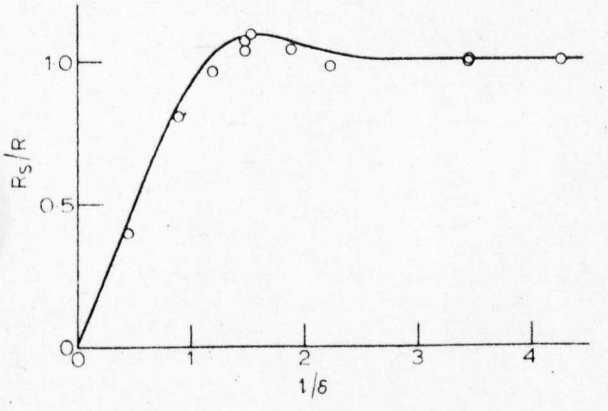


Fig. 2B *Ratio of surface resistances for a thick- and thin-film sample plotted against the ratio thickness/skin depth*

○○○ measured values
$R_s/R = 1$ corresponds to an evaporated copper film with $R_s = 0{\cdot}0286\,\Omega$, $\delta = 0{\cdot}725\,\mu$m, $\sigma = 48{\cdot}3 \times 10^6\,\Omega^{-1}\text{m}^{-1}$

Measurements using this technique yield a conductivity of $57{\cdot}1 \times 10^6\ \Omega^{-1}\,m^{-1}$ for electrolytically polished, vacuum-annealed high-purity bulk copper. This compares favourably with the standard d.c. conductivity for copper of $58{\cdot}0 \times 10^6\ \Omega^{-1}\,m^{-1}$.

Measurements on films with thickness l greater than three skin depths ($l/\delta > 3$) yield the bulk value of surface resistance R_s. For $l/\delta < 3$, the actual microwave surface resistance R for that thickness is given, since the intrinsic impedance of the insulating substrate greatly exceeds that of the film. R is related to R_s by the expression[2]

$$\frac{R_s}{R} = \frac{\cosh(2l/\delta) - \cos(2l/\delta)}{\sinh(2l/\delta) + \sin(2l/\delta)} \quad = (l/\delta \quad \text{for} \quad (l/\delta) < 0{\cdot}4$$

Measurements of R for evaporated copper films are compared with this theory in Fig. 2B, showing good agreement.

Further measurements on films deposited from various plating solutions will be published later.

Acknowledgment: The authors are grateful for support from the UK Science Research Council for this work.

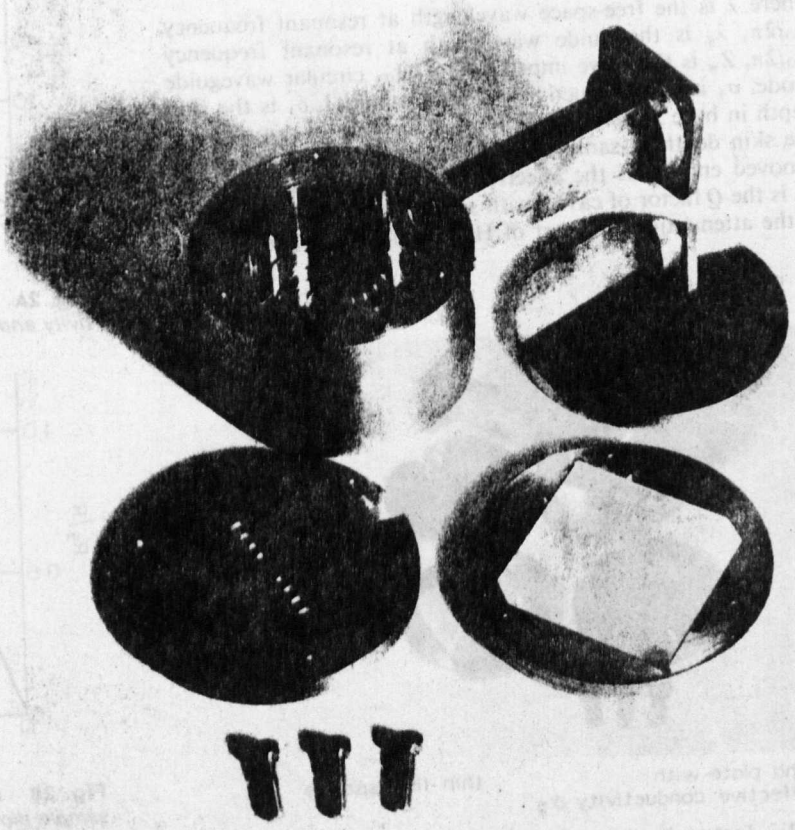
R. S. BUTLIN
M. K. McPHUN

Department of Engineering
University of Warwick
Coventry, War. CV4 7AL, England

30th November 1972

References

1 MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: 'Principles of microwave circuits' (McGraw–Hill, 1948), section 7.10
2 RAMO, S., WHINNERY, J. R. and VAN DUZER, T.: 'Fields and waves in communication electronics' (Wiley, 1965), section 5.19



APPENDIX 22.

Simple resonator method for measuring dispersion of microstrip.

Electronics Letters 8, 6, pp 165-166, (23rd March 1972).

SIMPLE RESONATOR METHOD FOR MEASURING DISPERSION OF MICROSTRIP

Indexing terms: Microwave measurements, Striplines, Permittivity measurement

The method uses the resonant frequencies of a simple straight microstrip resonator, short-circuited at both ends. It is excited by passing a current through one of the short circuits. Graphs of effective relative permittivity against frequency are given for resonators on alumina and quartz substrates, with an error of within only $\pm 0{\cdot}1\%$.

This letter presents a simple and practical microwave measuring technique, utilising a resonant line to determine the effective relative permittivity ε_{eff} of microstrip. Published works[1-7] in this field have shown that ε_{eff} increases with an increase in frequency, i.e. the line is dispersive. Hence the characteristic impedance of the line decreases with an increase in frequency. Therefore the variation of ε_{eff} with frequency should be known.

There are four methods of measuring ε_{eff} in common use:

(*a*) open-circuited resonant microstrip line[1]

(*b*) ring resonator[2, 3, 6]

(*c*) variation of phase of a line with frequency[3]

(*d*) nodal shift technique.[4]

Method *a* needs correction for the 'end effect', method *b* becomes difficult for wide strips of the microstrip line, method *c* requires correction for transition to the measuring equipment, and method *d* is very laborious. A comparison of these techniques[5] concluded that method *b*, the ring resonator, has been shown[6] to have a number of difficulties that require the use of a large substrate area to overcome. The new method presented overcomes these disadvantages.

Cooke[7] recognised that the short-circuit termination of a microstrip line could eliminate end effects. The method presented here uses a straight resonator with a short circuit at both ends. This configuration has resonant frequencies whenever the line length becomes a multiple of half a wavelength at that frequency.

The effective relative permittivity ε_{eff} for a microstrip line is defined such that

$$v = \frac{c}{\sqrt{\varepsilon_{eff}}}$$

where ε_{eff} is a function of frequency and the parameters of

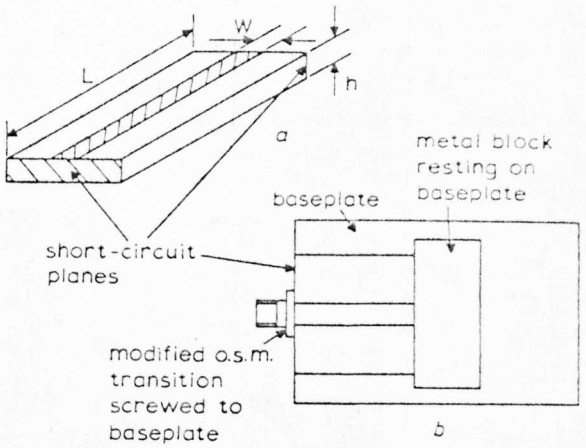


Fig. 1 *Microstrip resonator short circuited at both ends*
a Resonant line
b Test jig

the microstrip, and c is the velocity of light in free space. For an air-dielectric line with no dispersion, the resonant frequency f_0 would be

$$f_0 = \frac{nc}{2L}$$

where n is the number of halfwavelengths and L is the length of the microstrip line.

Let f_{ox} be the resonant frequency of the microstrip resonator with dispersion. Then

$$f_{ox} = \frac{nv}{2L} = \frac{nc}{2L\sqrt{\varepsilon_{eff}}}$$

Therefore

$$\varepsilon_{eff} = \left(\frac{nc}{2Lf_{ox}}\right)^2$$

Knowing the harmonic number n and the resonant frequency f_{ox} of the resonant line, ε_{eff} can be calculated.

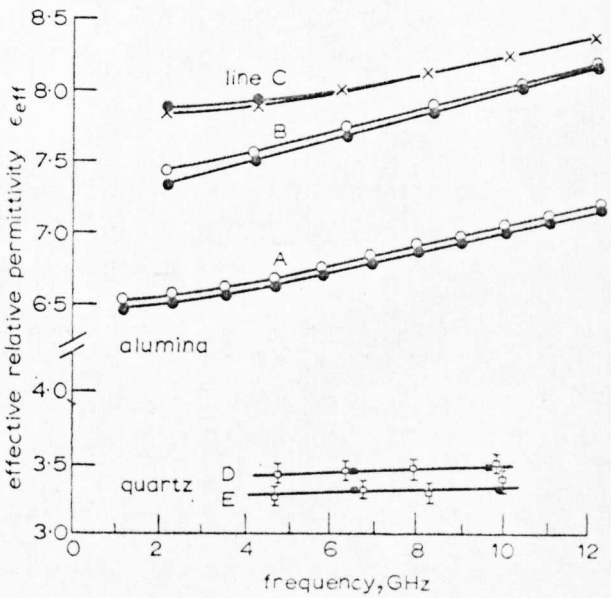


Fig. 2 *Variation of effective relative permittivity with frequency*
● ● ● with short-circuit block
○ ○ ○ without short-circuit block
× × × with silver-paint short circuits
□ □ □ nodal-shift technique
■ ■ ■ resonant-line technique

Table 1

Resonant line	A	B	C	D	E
Substrate material	alumina	alumina	alumina	fused quartz	fused quartz
Line length, mm	50·55	25·33	25·34	25·165	24·783
Strip width, mm	1·02	3·18	3·18	4·044	2·851
Substrate thickness, mm	1·02	1·02	0·635	0·489	0·499
Nominal impedance Z_0, Ω	50	26	20	18·5	24

Experimental work: Measurements of ε_{eff} against frequency were made in order to study the accuracy of this method. In particular, we examined the repeatability of the results and the effect of fringing fields on the short-circuit plane and the effect of using silver paint to make the short circuits. We also compared our results with those obtained by the nodal-shift technique.[4]

Measurements were made on the five lines shown in Table 1.

Short-circuit planes extended on both ends on the strip over the complete width of the ground plane, as shown in Fig. 1*a*.

The resonant line was excited through a single probe. The probe consisted of an o.s.m. microstrip adaptor, with the tab filed down to 0·3 mm by 0·2 mm, screwed onto the baseplate. The resonator was placed on the base plate, with the probe in contact with the short circuit to feed a current through it (Fig. 1*b*). This gave a sensibly constant coupling over the complete frequency range plotted. Careful checks were made to ensure that the coupling did not affect the resonant frequency. The measurements were made using an ordinary swept-frequency reflectometer connected to the probe. The resonance was indicated by a sharp decrease in the reflection coefficient.

If measurements of a restricted frequency range only are required, it is possible to find the value of n as follows: A hand-held probe (e.g. a screwdriver) is moved along the strip while the resonance is observed. The maximum shift of frequency is observed at every voltage maximum.

It was found that there is some 'field spillover' at the far end of the short-circuit plane. This can be overcome by a metal block placed as in Fig. 1*b* to extend the short-circuit plane to the fringing fields. At the probe end of the line, the body of the o.s.m. connector serves to extend the short circuit plane.

Fig. 2 shows some of the results obtained for ε_{eff} against frequency using this resonant-line technique. The variation of ε_{eff} for alumina substrates, with and without the metal block, is shown for lines A and B, and the effect of replacing the deposited short circuits with silver paint is shown for line C. For lines D and E, on quartz substrates, a comparison of results obtained by the resonant-line and nodal-shift techniques is shown. The latter gives a spread of ±2%, and excellent agreement between the two techniques was obtained. The expanded scale for ε_{eff} and the lack of scatter in the results are noteworthy.

It can be seen that failure to short-circuit the fringing fields can lead to errors of up to 1·4% for small n. The use of silver paint to form the short circuits does not affect the results too significantly. It does, however, lower the Q factor of the resonator and make the measurements much more difficult to perform.

The accuracy of this resonant-line method depends on the measurement of the length L of the line and the resonant frequency f_{ox}. L can readily be measured to a very high accuracy. Using a frequency counter, the accuracy of f_{ox} depends mainly on the Q factor of the resonator, typically 350. We have found that we can measure f_{ox} to an accuracy better than within ±0·1%. Thus, with care, a total accuracy of ε_{eff} of within ±0·1% can be achieved, and ±0·3% can be achieved with ease.

The method has advantages over other methods in its simplicity and accuracy, and it can be used on lines with wide strips where other methods become more difficult. For example, line D would require a 90 mm-square substrate to obtain accurate results using the ring-resonator method,[2, 6] whereas our results were obtained on a 25 mm-square substrate.

The authors wish to thank E. G. Coates, who supplied some of the samples, and the UK Science Research Council for financial support.

K. MEHMET
M. K. McPHUN
D. F. MICHIE

Department of Engineering
University of Warwick
Coventry CV4 7AL, England

21st February 1972

References

1 HARTWIG, C. P., MASSE, D., and PUCEL, R. A.: 'Frequency dependent behaviour of microstrip'. Proceedings of the GMTT Microwave Symposium, Detroit, 1968, pp. 110–116
2 TROUGHTON, P.: 'The evaluation of alumina substrates for use in microwave integrated circuits' *in* 1969 European Microwave Conference. *IEE Conf. Publ.* 58, pp. 49–56
3 CAULTON, M., *et al.*: 'Measurement of the properties of microwave integrated circuits'. Proceedings of the GMMT International Microwave Symposium, Dallas, 1969, Texas, pp. 38–44
4 SECKELMANN, R.: 'Nodal shift measurement to determine transmission line properties', *Microwave J.*, 1968, pp. 69–72
5 ARNOLD, S.: 'Dispersion effects in microstrip on alumina substrates', *Electron. Lett.*, 1969, **5**, pp. 673–674
6 WOLFF, I., and KNOPPIK, N.: 'Microstrip ring resonator and dispersion measurement on microstrip lines', *ibid.*, 1971, **7**, pp. 779–781
7 COOKE, R. E.: 'Dispersion characteristics of microstrip transmission line' 1969 *in* European Microwave Conference. *IEE Conf. Publ.* 58, p. 2

APPENDIX 23.

Calibration of microwave network analyser for computer-corrected s parameter measurements.

Electronics Letters 9, 6, pp 126-128, (22nd March 1973).

ELECTRONICS LETTERS 22nd March 1973 Vol. 9 No. 6 pp. 126–128

CALIBRATION OF MICROWAVE NETWORK ANALYSER FOR COMPUTER-CORRECTED S PARAMETER MEASUREMENTS

Indexing terms: Network analysers, Calibration, Microwave measurements, Electronics applications of computers

A new method of calibration that uses three short-circuits (one direct and two offset short-circuits) enables the scattering parameters of the error network to be evaluated, and makes corrected measurements of the reflection coefficients possible without requiring the use of matched terminations. The theoretical results are confirmed by experimental results.

Most microwave network analysers[1] measure the phase and amplitude of the forward and reflected waves to obtain the reflection coefficient. These waves are normally extracted by directional couplers, and downconverted to an intermediate frequency, where easier phase and amplitude comparisons may be made. Errors result chiefly from directivity effects in the couplers, tracking and crosscoupling in the forward and reflected measuring channels, and the mismatch between the connectors and the signal source.

Computer correction[1, 2] may be applied to calibrate out these errors. However, in these correction systems, a short-circuit, an offset short-circuit, and a well calibrated or sliding termination are required. A method proposed by Shurmer[3] obviates the use of a sliding matched termination and is of particular use where space is at a premium. However, at higher frequencies, Shurmer's method requires the capacitance of an open-circuit line to be determined, but this is avoided in the procedure presented here. For measurements under many conditions, the sliding termination causes most difficulty. Such terminations are not generally available, except for airspaced coaxial lines and waveguides. Also, there may be insufficient room to accommodate a sliding termination, as is the case with the author's present work on a jig for measuring lumped inductors at X band frequencies. Thus, hitherto, it has been impossible to apply computer correction right up to the desired measurement plane under many circumstances. This undesirable state of affairs may be resolved by the use of an additional offset short-circuit to replace the sliding matched termination.

The errors in a network-analyser system may be represented[1] by assuming a perfect network analyser and a 2-port error network interposed between it and the device under test. The error network may be conveniently characterised in terms of its scattering parameters, and will assume the form shown in Fig. 1. For calibration purposes, the test device is removed, and the error network is terminated in turn by the three short-circuits shown in Fig. 1. From basic theory,

$$\rho_1 = S_{11} - \frac{S_{21}S_{12}}{1+S_{22}} \quad \ldots\ldots\ldots \quad (1)$$

$$\rho_2 = S_{11} - \frac{S_{21}S_{12}e^{-j2\phi}}{1+S_{22}e^{-j2\phi}} \quad \ldots\ldots\ldots \quad (2)$$

$$\rho_3 = S_{11} - \frac{S_{21}S_{12}e^{-j2(\phi+\theta)}}{1+S_{22}e^{-j2(\phi+\theta)}} \quad \ldots\ldots\ldots \quad (3)$$

From eqns. 1 and 2, eqn. 4 may be obtained, and, similarly, eqn. 5 is derived from eqns. 2 and 3:

$$S_{11}(e^{-j(2\phi)}-1) = S_{22}e^{-j2\phi}(\rho_1-\rho_2)+\rho_1 e^{-j2\phi}-\rho_2 \quad \ldots \quad (4)$$

$$S_{11}(e^{+j2\theta}-1) = S_{22}e^{-j2\phi}(\rho_3-\rho_2)+\rho_3 e^{+j2\theta}-\rho_2 \quad \ldots \quad (5)$$

Solution of eqns. 4 and 5 yields

$$S_{11} = \frac{\rho_1\rho_2(e^{-j2\phi}-1)-\rho_2\rho_3(e^{+j2\theta}-1)-\rho_1\rho_3(e^{-j2\phi}-e^{+j2\theta})}{(e^{-j2\phi}-1)(\rho_2-\rho_3)-(e^{+j2\theta}-1)(\rho_2-\rho_1)} \quad \ldots\ldots \quad (6)$$

Eqns. 1 and 2 also give

$$S_{22} = \frac{e^{+j2\phi}(\rho_2-S_{11})+S_{11}-\rho_1}{(\rho_1-\rho_2)} \quad \ldots\ldots \quad (7)$$

Eqn. 1 gives

$$S_{21}S_{12} = (S_{11}-\rho_1)(1+S_{22}) \quad \ldots\ldots\ldots \quad (8)$$

For the device under test, with a reflection coefficient Γ_l,

$$\rho = S_{11} + \frac{S_{21}S_{11}\Gamma_l}{1-S_{22}\Gamma_l} \quad \ldots\ldots\ldots \quad (9)$$

from which

$$\Gamma_l = \frac{\rho - S_{11}}{S_{22}\rho + S_{21}S_{12} - S_{11}S_{22}} \quad \ldots\ldots\ldots \quad (10)$$

Thus, with the substitution of S_{11}, $S_{21}S_{12}$, and S_{22}, from eqns. 6–8, into eqn. 10, the value of Γ_l can be obtained.

Owing to the complexity of the equations, attempts to assess the possible errors in the parameters of the error network using the theory of errors were not practicable. However, stringent conditions of measurements were imposed by choosing a test device whose reflection coefficient was of the same order as that of the error network. The measurements shown in Table 1 were made manually, whereby the 'unknown' (a coaxial matched load with an APC 7 connector) was measured through three different mismatches. In each case, the calibration was carried out on the 'unknown' side of the mismatch. When the results are compared on a

Table 1

Conditions of measurements	Measured reflection coefficient			Frequency	Calculated errors			Device under test	
	ρ_1	ρ_2	ρ_3		S_{11}	$S_{21}S_{12}$	S_{22}	Measured ρ	Corrected Γ_l
				GHz					
Through instrument*	1	1	1	1·01143	0·022	0·999	0·022	0·025	0·0199
	∠−180	∠−70	∠43		∠−18·24	∠−0·77	∠−162·5	∠−68	∠−124·7
Through instrument* and pair of o.s.m. connectors	1	0·98	1	1·01215	0·020	0·994	0·00918	0·006	0·021
	∠−180	∠−69	∠48		∠89·7	∠1·5	∠132	∠−178	∠−108·4
Through instrument* and pair of GR 874 connectors	1	1	1	1·01220	0·0187	0·999	0·0187	0·005	0·022
	∠−180	∠−68	∠47		∠44·14	∠1·47	∠137·3	∠−92	∠−128·5

* Hewlett-Packard 8746B

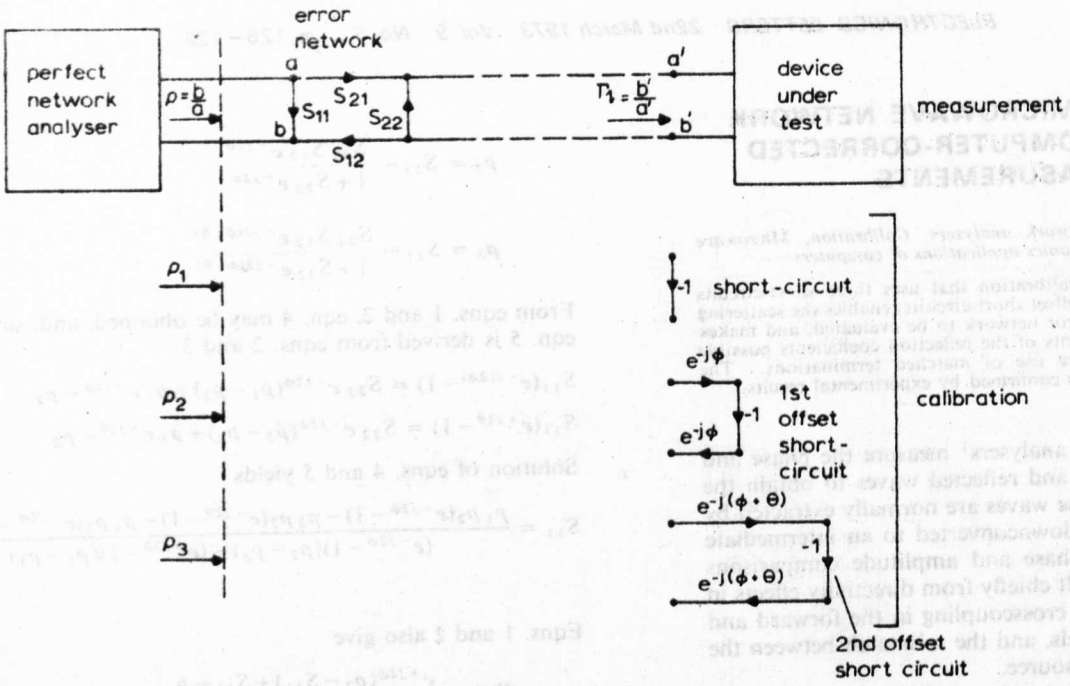


Fig. 1 *Signal-flow graphs of system under calibration*

Smith chart, it is seen that the maximum variations in Γ_t can be contained within a circle corresponding to $|\Gamma| = 0{\cdot}0035$. Most of this small variation is in the phase of Γ_t, and we believe this results from variations in the frequency during measurement.

Thus we conclude that this is a powerful new technique for enabling computer correction to be applied at many measurement planes not possible before. In addition to not requiring a matched termination, the additional offset short-circuit permits more accurate phase measurements to be read, as its reflection coefficient is nearer the periphery of the Smith chart. The disadvantage with this method is that care must be exercised in choosing the offset lengths, so that the denominator of eqn. 6 does not vanish to zero.

E. F. DA SILVA
M. K. MCPHUN

Department of Engineering Science
University of Warwick
Coventry, War. CV4 7AL, England

16th February 1973

References

1 HAND, B. P.: 'Developing accuracy specifications for automatic network analyser systems', *Hewlett-Packard J.*, Feb. 1970, p. 16
2 HACKBORN, R. A.: 'An automatic network analyser system', *Microwave J.*, 1968, **11**, (5), pp. 45–52
3 SHURMER, H. V.: 'Calibration procedure for computer-corrected *S* parameter characterisation of devices mounted in microstrip' (submitted to *Electron. Lett.*)

APPENDIX 24.

Measurement of overlay capacitors at X-band with independent assessment of thin-film conductor losses.

Proceedings of the 1973 European Microwave Conference, Brussels, September 1973, paper B. 14. 5.

MEASUREMENT OF OVERLAY CAPACITORS AT X-BAND WITH INDEPENDENT ASSESSMENT OF THIN-FILM CONDUCTOR LOSSES

R.S. BUTLIN, D.F. MICHIE AND M.K. McPHUN

Department of Engineering,
University of Warwick,
Coventry,
CV4 7AL.
England.

The accurate characterization of thin film microwave overlay capacitors has been impossible in the past, due to transition errors introduced by the enormous difference in size between the components and the measurement system. Most errors, particularly of loss, are due to the use of bond wires.

In this new technique the test capicitator is fabricated as an intrinsic part of an X-band microstrip resonator, eliminating the bond wires, and the capacitor can be measured in the same form as would be used in a practical circuit. Figs. 1 and 2 show the resonator. It is $3\lambda/4$ long with one end short circuited to the ground plane, and the other having a small gap coupling to the measurement system. The test capacitor is in series with the transmission line, positioned at a current maximum.

Conventional thin film techniques are used to construct the resonator. The test capacitor is constructed across a narrow gap in the resonator in two stages. At first it is made without a top electrode, and the resonant frequency and Q of the resonator plus capacitor are measured. Then the top electrode is completed and the measurements are repeated. The capacitor Q factor and capacitance are determined from these measurements and the physical dimensions of the resonator. The capacitance can be measured to within ± 1.2% and the Q factor to within ± 7%, using a resonator with an unloaded Q of 400, deposited on a quartz substrate.

In Fig. 1c is shown a photograph of a completed 1 pF capacitor taken using a scanning electron microscope. Typically the lower electrode consists of 0.75 μm of evaporated copper, the dielectric is 1 μm thick SiO_2 deposited by R.F. sputtering, and the 5 μm thick copper upper electrode is first evaporated and then electroplated. A new photolithographic technique[1] was developed to make these capacitors.

The value of capacitance measured depends on two factors: the true capacitance and the electrode inductance. Similarly the Q factor depends on the electrode loss and the dielectric loss. Short electrodes, i.e. with width/length ratio $W/L > 4$, render the inductance negligible. To separate the contributions to the loss of the capacitor, independent measurements of dielectric loss[2] and electrode conductivity are needed. Here we report our measurements of electrode conductivity.

The Authors wish to thank the U.K. Science Research Council for financial support.

Experience shows that the surface resistance of the films used cannot be predicted from d.c. or l.f. measurements. A new cavity measurement technique[3] is used to measure the surface resistance in X-band. The sample metal film is deposited on a substrate and forms one end plate of an H_{011} cavity. Measurement of the return loss at resonance alone permits the sample surface resistance to be obtained directly from a graph. The maximum error in surface resistance is ± 0.0012 ohms.

Results one evaporated copper films[3] show that the microwave conductivity does not depend on film thickness, and the theoretical relation between film thickness and surface resistance is followed closely in practise. A series of factorially designed experiments[4] was made on films deposited from acid copper electroplating solutions. The effects of plating parameters were related to surface finish, surface resistance, and d.c. conductivity of the films. The composition of the plating solution was shown to be the most important factor.

Table 1 summarizes results obtained at 10 GHz from films deposited using evaporation and electroplating solutions. Two films were deposited from each solution and their microwave surface resistance measured; the d.c. conductivity of one of the films was then measured. The other film was annealed at 250°C for 2 hours in vacuum, and its microwave surface resistance and d.c. conductivity then measured. The copper pyrophoshate solution initially gave the best result, but these solutions deteriorate rapidly, and this result could not be repeated.

We conclude that lowest surface resistance is obtained from acid copper solutions with added brightener, and the effect of annealing on these films is small. Measured microwave conductivities are between 89% and 94% of the measured d.c. values, an effect we attribute to impurities in the film, not surface roughness. Use of silver plating is not advised. Annealing of plated gold films is required to achieve low loss.

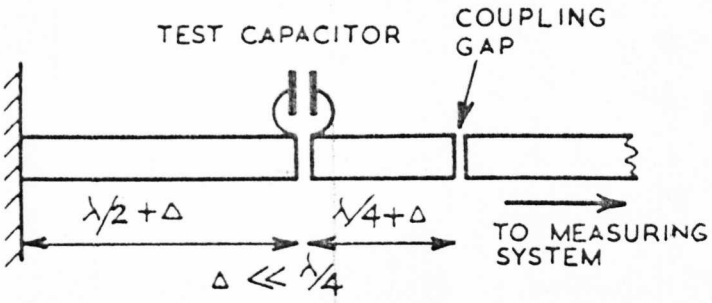
Some results of measurements on capacitors are shown in table 2.

We conclude that for these dielectrics, the dielectric loss predominates for capacitors with values smaller than 10 pF, the electrode loss becoming significant for C > 10 pF. The electrode loss is dominated by the lower electrode whose thickness is limited by the need for the dielectric to overlap it. Decreasing the lower electrode thickness below 0.75δ (δ=skin depth) gives a rapidly increasing loss. If the upper electrode thickness exceeds 3δ, its contribution to the electrode loss is negligible. The electrode Q factor can be greatly increased by using the configuration of Fig. 3b rather than Fig. 3a. This results in an increase of electrode Q of 75% for W/L = 1.

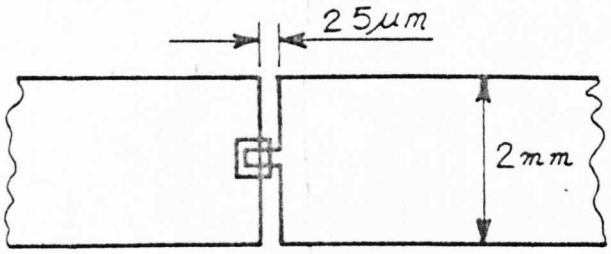
We expect R.F. sputtered alumina to give much higher dielectric Q factors when the deposition process is optimized.

References

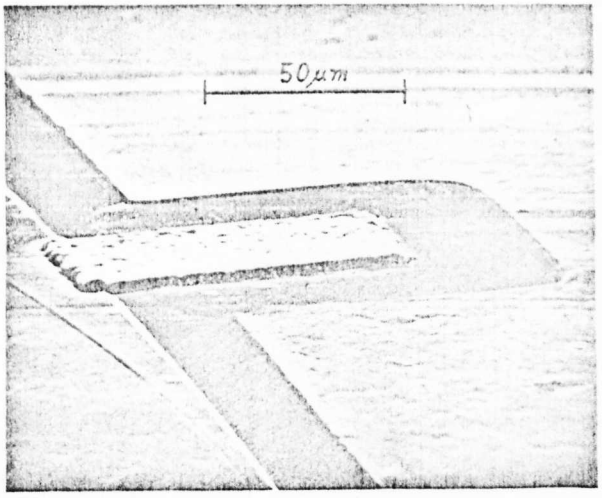
1. R.S. Butlin and D.F. Michie "Method for depositing material on selected areas of a surface" U.K. Patent Application 5342/72.
2. M.K. McPhun and K. Mehmet "Thin film dielectric measurements" chapter of "High frequency dielectric measurements" IPC Science and Technology Press 1973.
3. R.S. Butlin and M.K. McPhun "Surface-resistance measurements of thin conducting films at 10 GHz" Electronics Letters 8, 26, p 637-639, 28th Dec. 1972.
4. O.L. Davis "Design and analysis of industrial experiments" Oliver and Boyd, 1954.



a). SCHEMATIC



b) CAPACITOR MOUNTED IN GAP



c) DETAIL OF CAPACITOR

FIG. 1. MICROSTRIP RESONATOR WITH OVERLAY CAPACITOR

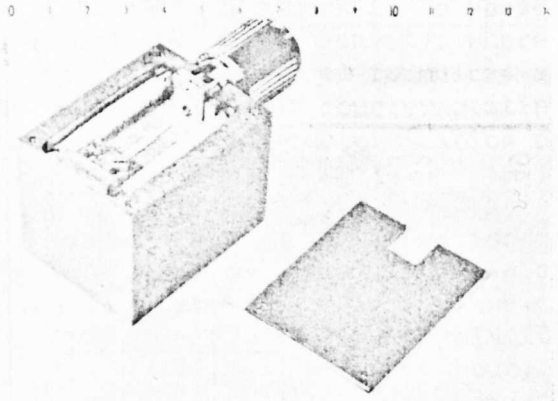


FIG.2. MICROSTRIP RESONATOR MOUNTED IN SHIELDING BOX.

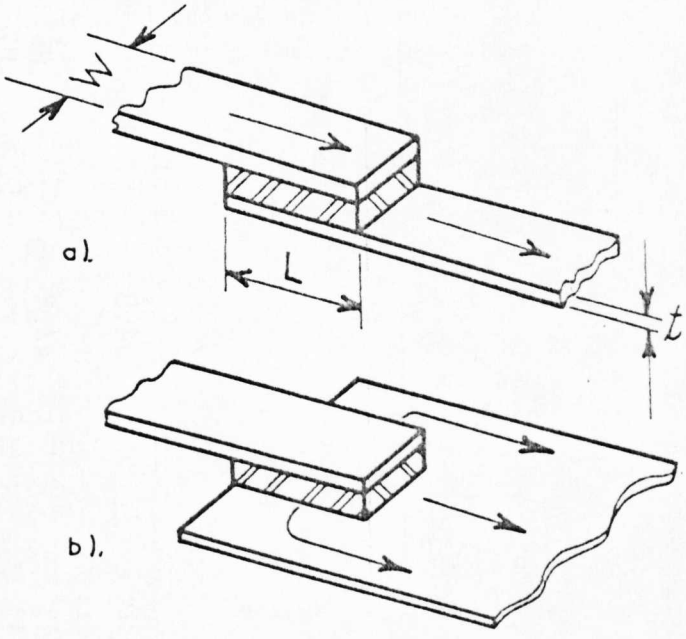


FIG. 3. TWO CAPACITOR CONFIGURATIONS (ARROWS INDICATE DIRECTIONS OF CURRENT FLOW.)

Dielectric	C(pF)	Q	Q electrode	Q dielectric	t(μm)	W/L
SiO_2	1.1	122	1557	133	1.1	1.6
SiO_2	1.0	120	1602	129	0.91	2.01
SiO_2	0,71	137	2349	145	0.72	3.0
Al_2O_3	1.25	164	1803	178	1.1	2.67

TABLE 2 Measurements on Capacitors at 9.5 GHz

TABLE 1 Measurements on metal films at d.c. amd 10 GHz

Deposition Method	Thickness (μm)	Surface resistance (ohms)	annealed	Microwave Conductivity $(\Omega m)^{-1}$	annealed	DC Conductivity $(\Omega m)^{-1}$	annealed	CLA (μm)	Theoretical DC Conductivity $(\Omega m)^{-1}$	Measured/Theoretical DC Conductivity (%)	annealed	Measured/Theoretical Microwave Conductivity (%)	annealed
		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{6}$	$\times 10^{6}$	$\times 10^{6}$	$\times 10^{6}$		$\times 10^{6}$				
P.M.D. gold	5.1	3.38		34.5		36.9		0.07	41.0	90		84.1	
	4.6	3.38	3.28	34.5	36.6		40.4	0.07			98.6	84.1	89.3
Englehardt E56 gold	2.4	3.34		35.2		38.1		0.03	41.0	92.9		86.0	
	2.4	3.33		35.3		37.6		0.03	41.0	91.8		86.2	
Evaporated gold	2.5	3.48		32.5		34.8		0.02	41.0	84.9		79.3	
P.M.D. silver	3.3	3.03		43.3		47.8		0.05	62.9	76.0		68.9	
	3.3	3.00	2.91	43.7	46.5		52.0	0.07			82.6	69.5	73.9
Acid copper	3.5	2.98		44.2		49.7		0.05	58.0	84.6		76.2	
	3.5	2.99	2.96	44.0	44.9		50.5	0.04			87.1	75.9	77.5
Bright acid copper	3.2	2.91		46.6		52.7		0.03	58.0	90.8		80.4	
	3.2	2.90	2.84	46.8	48.8		53.0	0.03			91.3	80.7	84.2
Copper pyrophosphate													
Best result	4.5	2.89		47.1		47.2		0.03	58.0	81.4		81.2	
Worst result	4.8	3.05		42.3		45.4		0.04	58.0	78.4		73.0	
	4.8	3.05	3.03	42.3	42.9		46.0	0.04	58.0		78.3	73.0	73.9
Evaporated copper	3.1	2.86		48.1		52.1		0.01	58.0	89.9		82.9	

APPENDIX 25.

Effect of parasitics on the design of overlay capacitors in microstrip circuits.

I.E.E. Colloquium on Microwave Integrated Circuits, London, 24th May 1974.

THE EFFECT OF PARASITICS ON THE DESIGN OF OVERLAY CAPACITORS IN MICROSTRIP CIRCUITS.

D. Michie and M.K. McPhun.

The accurate characterization of thin film microwave overlay capacitors has been impossible until recently, due to transition errors introduced by the enormous difference in size between the components and the measurement system, and the effect of bond wires. Recently introduced measurement techniques have overcome these difficulties by fabricating the test capacitor as part of a microstrip resonator.

Measurements on capacitors using this technique in X-band have shown that parasitic electrode inductance and resistance have considerably larger effects than is generally appreciated. They should be made as small as possible.

Measured results at 8.6 GHz (Fig. 1) show that the inductive reactance due to the electrodes was up to 9% of the reactance due to the true capacitance.

Theoretical analysis of the overlay capacitor has been based on a transmission line model. The surface resistance of the electrodes was separately assessed for insertion in the model.

Electrode Resistance The cross-section of the capacitor in Fig. 2 shows that the lower electrode must be thin; it therefore contributes most to the electrode loss. The optimum lower electrode thickness is 1 skin-depth. The upper electrode thickness should be at least 1.6 skin depths. The lower electrode losses are then 20% higher than the upper electrode losses.

Copper material gives an electrode Q up to 30% higher than does gold. Alternatively a copper lower electrode can be half the thickness of a gold one for the same losses. This allows the use of thinner dielectric layers.

The electrode Q as a function of capacitor aspect ratio (width/length) and capacitance at 10 GHz is shown in Fig. 3. Electrode loss predominates for capacitance greater than 25 pF due to the practical limit on aspect ratio. For low aspect ratios the configuration of Fig. 4a) is much better than that of Fig. 4b).

Electrode Inductance is a function of capacitor width and length, which should be as short as possible. Again a large aspect ratio is required. Fig. 5 shows that for electrode inductive reactance less than 10% of the capacitive reactance, aspect ratios of 10 for a 1 pF capacitor, and 200 for a 10 pF capacitor are required for the configuration of Fig. 4b). Larger capacitances are allowable with the configuration of Fig. 4a).

The inductance of microstrip feeders can be predominant. Great care is required in evaluating their contribution in terms of the particular circuit application.

The authors wish to thank the Science Research Council for financial support.

References BUTLIN, R.S.; MICHIE, D;McPHUN, M.K. "Measurement of overlay capacitors at X-band with independent assessment of thin film conductor losses" Proc. 1973 European Microwave Conference, Brussels, Paper No. B.14.5.

MICHIE, D. Ph.D. Thesis, University of Warwick, 1974.

Department of Engineering, University of Warwick, Coventry.

DIELECTRIC MATERIAL	MEASURED C (pF)	TRUE C (pF)	$\frac{X_L}{X_C}$ (%)	UPPER FEEDER LENGTH (μm)	ASPECT RATIO	ELECTRODE Q	DIELECTRIC Q	MEASURED Q
SiO_2	0.75	0.73	2.9	22	1.7	2490	161	147
SiO_2	0.70	0.66	5.5	47	3.0	2280	172	151
SiO_2	1.02	0.93	8.2	55	2.4	1945	174	148
Al_2O_3	1.26	1.14	9.4	50	2.7	1465	240	189

FIG. 1. MEASUREMENTS AT 8.6 GHZ.

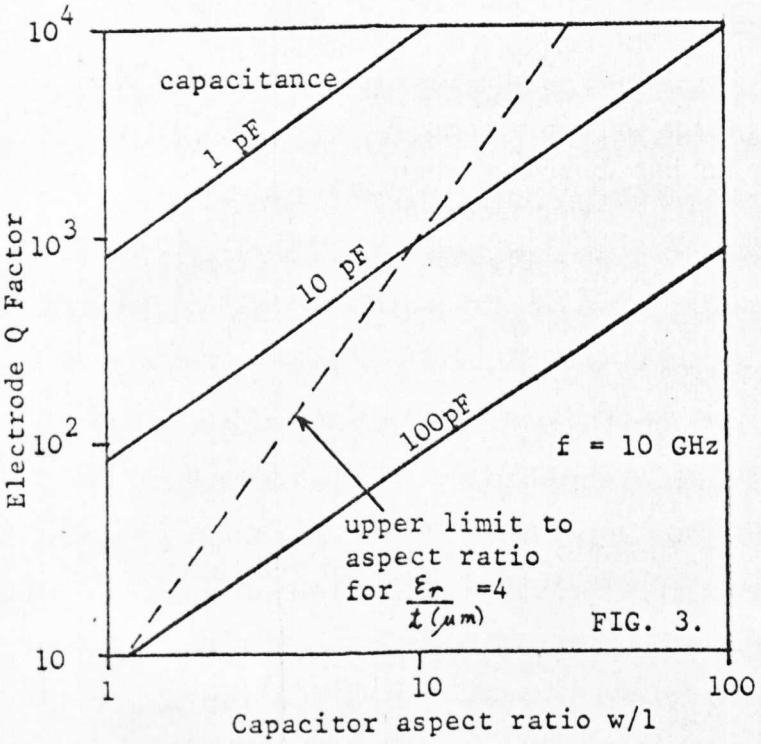


FIG. 3.

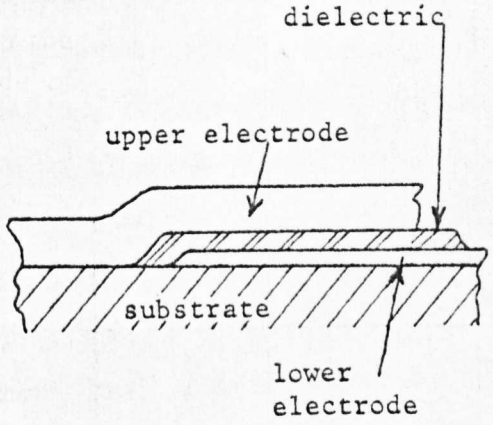


FIG. 2.

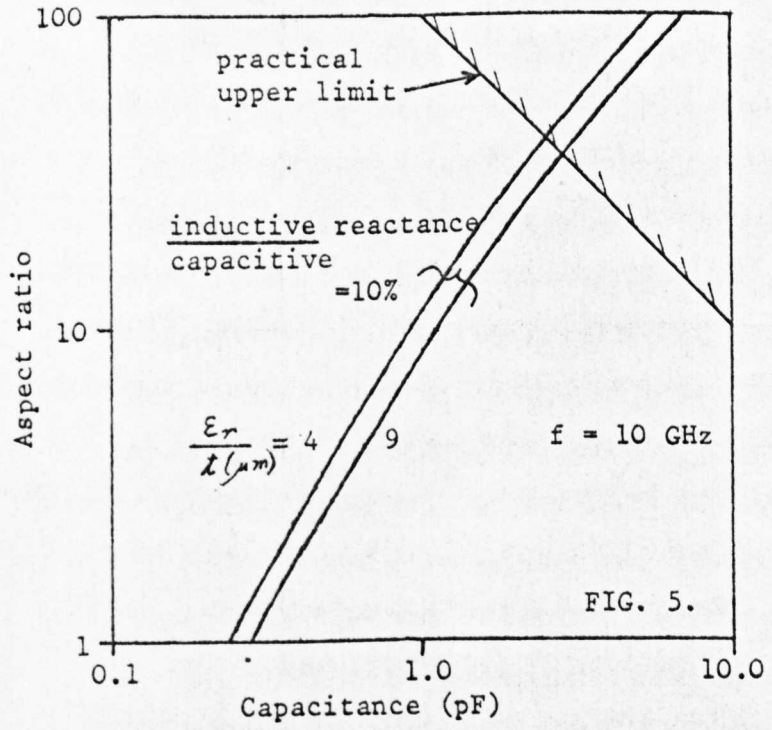


FIG. 5.

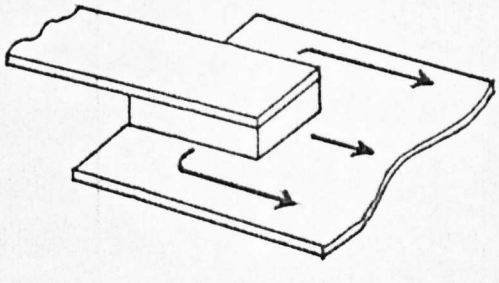


FIG. 4a).

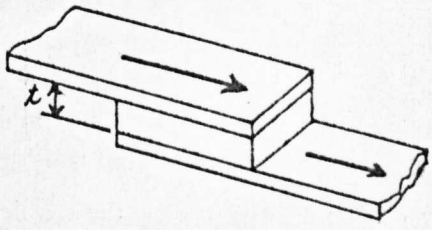


FIG. 4b).