THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 31 APRIL 1966 NUMBER 4

TECHNICAL INNOVATION AND GOVERNMENT

INDUSTRIAL development and international commercial competitiveness have now reached the stage at which there must be concerted and co-ordinated efforts by governments to encourage the rapid exploitation of new technological developments by private firms. Economic growth can only come from technical innovation which is essentially the introduction of new or improved products or processes through the exploitation of scientific and technical knowledge.

The preparatory work by the Interim Committee for the recent Ministerial Meeting on Science held by the O.E.C.D. countries put forward some very cogent thinking on the ways in which governments can stimulate and encourage technical innovation. Co-operative research institutes and technical information and advisory services already have an important function in certain—and especially old-established—industries. Again, some European Governments are taking measures to strengthen technical information and advisory services for small firms in industry.

In the view of the Committee, however, positive government support for civil technological innovation in specific industrial sectors will become increasingly important where the required scale of R & D effort is large. Civil development contracts, deliberate policies for government procurement of technologically advanced products, and measures to promote industrial concentration and the transfer of scientific and technical 'know-how' from the defence/space sector to the civil sectors, are suggested as being ways whereby governments can foster innovations.

All these courses of action imply a greater awareness and competence in government concerning the problems of technological innovation. National bodies responsible for science policy have a vital rôle to play, and in some countries governments are discussing whether to create agencies specifically responsible for stimulating innovation in industry. In the United Kingdom the Ministry of Technology was established for this purpose in 1964.

The feasibility of such measures will vary considerably amongst different countries, depending on their level of technological development and on the size of their population and their stock of qualified scientists and engineers. Thus, in developing countries, particular—but not exclusive—emphasis may need to be put on R & D undertaken in order to be able to identify the foreign technology which it would be advantageous to import and adapt to local conditions. Even the industrially advanced European countries import a considerable amount of foreign technology, notably from the United States, with which they have a growing deficit in payments for patents and manufacturing licences.

A first major step in solving these problems, according to the Interim Committee, would be to recognize that the problem is essentially one of the size of the R & D effort, and of the market, required for the development of certain advanced technologies. Most European countries are not big enough to make important technological contributions in all these sectors and many of them have therefore recognized the importance of arriving at some degree of concentration of national R & D effort in these sectors, and of exploring possibilities for international co-operation in technological development at both the firm and the governmental level.

An important stage in this co-operation would be to exchange experience on the use of specific measures for stimulating technical innovation, particularly civil development contracts, government procurement, and on technical information and advisory services for industry. Great Britain will certainly welcome this co-operation with Europe and elsewhere, and for its part can speak of extensive experience in many ways of encouraging innovation.

INSTITUTION NOTICES

Payment of Subscriptions in Local Currencies

Overseas members and subscribers are reminded that they may make subscription and other payments in local currency to the account of the Institution at the following Banks:

AUSTRALIA:	Bank of New South Wales, Head Office, Sydney.
CANADA & U.S.A.:	Canadian Imperial Bank of Commerce, 119 Sparks Street, Ottawa, or King and Victoria Branch, Toronto, 1.
FRANCE:	Barclays Bank Ltd., 33 Rue du Quatre Septembre, Paris 2.
INDIA:	State Bank of India, P.O. Box 10, Bangalore 1.
ISRAEL:	Barclays Bank D.C.O., 103 Allenby Road, Tel-Aviv.
NEW ZEALAND:	Bank of New South Wales, Auckland.
PAKISTAN:	National Bank of Pakistan, P.M.A. Building, Nicol Road, Karachi 2.
SOUTH AFRICA:	Barclays Bank D.C.O., 40 Simmonds Street, Johannesburg.

Advice of such remittances should be sent at the same time to the Institution's head office in London or to the Divisional Offices in Canada or India.

Conference on Steerable Aerials

The Conference on 'The Design and Construction of Large Steerable Aerials for Satellite Communication, Radio Astronomy and Radar' which is being sponsored by the Electronics Division of the Institution of Electrical Engineers, the Institution of Electronic and Radio Engineers, the Institution of Mechanical Engineers, the Institution of Structural Engineers and the Institute of Electrical and Electronics Engineers, will be held at the Institution of Electrical Engineers, Savoy Place, London, W.C.2, from 6th to 8th June 1966.

Six main sessions will cover electrical design (including profile errors, feeds, performance measurements and radomes) and mechanical design (including mounts and drives, structure and measurements). The final session will deal with control and tracking techniques.

A programme of technical visits to Goonhilly, Jodrell Bank and R.R.E. Malvern is being arranged, and there will be a cocktail party and dinner.

The fees for the Conference are £7 for members of the sponsoring societies and £9 for non-members. Registration forms and further information may be obtained from I.E.R.E., 8–9 Bedford Square, London, W.C.1.

I.R.E.E. Australia Convention

The Institution of Radio and Electronics Engineers Australia has announced that its next Radio and Electronics Engineering Convention is to be held in Sydney from 22nd to 26th May, 1967.

Original papers within the following broad subject areas are invited for the Convention:

Basic sciences and techniques; industry and industrial electronics; communications; electronic systems; computers and data processing; instrumentation; materials, components and production processes; bio-medical electronics; and professional activities (including educational aspects).

The deadline date for submission of manuscripts of completed papers is 1st December 1966. Offers of papers should be sent to the General Secretary, The Institution of Radio and Electronics Engineers Australia, Box 3120 G.P.O., Sydney, New South Wales, from whom further information may be obtained.

Commonwealth Telecommunications Conference

A Commonwealth Conference was held during March in London at which representatives met to discuss the setting-up of a new organization to strengthen Commonwealth co-operation throughout the whole field of external telecommunications. The Conference, which met at Marlborough House, was opened by the Postmaster-General, The Right Honourable Anthony Wedgwood Benn, and was closed on 22nd March by the Director-General of the Post Office, Sir Ronald German, K.C.B., C.M.G.

It is understood that the Conference has also made recommendations to Commonwealth countries for modifications to the existing arrangements between the countries for sharing the costs of their external telecommunications.

On 17th March many of the delegates attended a reception at 9 Bedford Square, where they were able to meet senior members of the Institution. The reception thus provided a valuable opportunity for exchange of views on engineering matters. The guests and members were received by the President, Colonel G. W. Raby, C.B.E.

Proceedings of Symposium on Microwave Applications of Semiconductors

The final Proceedings of the above Symposium, which was held at University College, London, in July 1965, has now been published. It comprises the full text of all 40 papers as well as reports of discussions, and may be obtained from the Institution at £6 per copy.

U.D.C. 028.8: 621.37/9

Controlling the Information Explosion

By

ADMIRAL OF THE FLEET THE

EARL MOUNTBATTEN OF BURMA,

K.G., P.C., O.M., D.Sc., C.Eng.

(*Honorary Member*)†

A Friday Evening Discourse given at the Royal Institution of Great Britain, Albemarle Street, London, on 28th January 1966.

Summary: The Selective Dissemination of Information or S.D.I. system is aimed at bringing to the individual scientist's attention only those new items of information that are of interest to him. This is done by comparing his subject interests in an electronic digital computer with the subjects covered by each document entering the information system, and printing out from the computer only those items which sufficiently match his interests. In this way a personal service can be provided to a large number of scientists, with feedback allowing progressive tailoring of the service to the needs of each individual.

The National Electronics Research Council is at present engaged in the first stage of a three-year investigation of the value, economics and acceptability of the S.D.I. system for electronics research workers. Although the investigation will initially be confined to the subject of electronics research, the system developed will be immediately applicable to any other branch of science and technology.

1. The Information Explosion

The 'Information Explosion' is a much-quoted phrase whose currency may have been depreciated with over-use; but it does highlight a very real problem. Of course the information problem has existed for a very long time. The Preacher's complaint in *Ecclesiastes*, 'Of making many books there is no end . . .' shows that we do not face a new problem. But the problem we face, although similar in kind, is quite different in degree. It is the increasingly rapid growth in numbers of documents particularly articles in periodicals and research reports, that give us good reason to refer to it as an explosion.

At the beginning of the nineteenth century the number of scientific journals and periodicals was about 100; in 1850 it was 1000; and by 1900 it had reached 10 000. Some estimates of the number at the present time go as high as 100 000.‡ If this rate of growth remains constant it would give a figure in the neighbourhood of one million at the end of this century.

Of course, extrapolations such as these are of little value. They rest on doubtful foundations and do not take into account other factors which could be expected to reduce the rate of growth. But I hope what I have said has convinced you that there is a real problem in dealing with the growing mass of information, which is growing at an ever-increasing rate.

2. The Current-Awareness Problem

The information problem is divided into two parts: Information Retrieval and Current Awareness. The part that is most often discussed is Information Retrieval. In this, one is seeking to find what has already been written on a subject or to find an answer to a specific question—what is commonly called 'retrospective searching'.

I do not for one moment wish to suggest that this is not a difficult problem. One has only to think of the number of books and periodicals in the British Museum, or in the National Lending Library of Science and Technology at Boston Spa to realize that it is not going to be easy to lay one's hand quickly on the book or article that will provide the answer required. However, the problem has been with us for a very long time—in fact since the first library or collection of information was set up—and at least a partial solution has been found. There is a large apparatus of reference works, bibliographies, abstract journals and classifications which allow any inquirer to find almost any information he requires—although one must admit it may take him a very long time to complete his search.

But the point is that there is nothing in the nature of retrospective searching, or Information Retrieval, which prevents the inquirer spending as long as required on the search. He himself may not have the time to spare but that is a different matter.

With the other type of information problem, keeping abreast of the literature, or 'Current Awareness', the situation is very different. Here the individual wishes to ensure he misses nothing of interest to him that is published in any of the enormous number of books, periodicals and reports which issue regularly from the printing presses or duplicating machines of the world.

I think it would help to visualize the problem by picturing a conveyor belt, to represent all the

† Chairman, National Electronics Research Council, 50 Bloomsbury Street, London, W.C.1.

‡ Pierre Auger, ECOSOC E/3362 (13th May 1960).

organizations publishing information. The belt is a very wide one and on it are books, periodicals and reports; as they come out and in any position. The searcher is on a platform just above the belt and as the information material passes underneath he can pick up and read anything that he thinks might be of interest to him. You can imagine his frustration as he realizes that for every item he takes time to examine, hundreds of others of possible interest to him have passed by.

This may seem an overdrawn picture but so far from being that, it is inadequate in two important particulars. Our friend on the platform has two advantages that no-one in real life has: he has immediate access to all the information and, of more importance, he can at least know of the existence of all the information that might be of interest, if he does not stop to read any publication.

In practice, in spite of the proliferation of different aids to current awareness (many of which I admit are extremely useful) it is virtually impossible for any individual, let alone someone as busy as a research worker, to read all the actual books, periodicals and reports, or even the current awareness publications listing them, in order to discover information of interest to him.

It certainly seems that it is in current awareness that the information explosion has got completely out of hand. It would be no solution to publish another announcement bulletin, since this would merely add to the material to be searched regularly by each individual.

A theoretical solution would be for each person to have his own agent permanently stationed in a large library, such as the British Museum. The agent would have to scan each piece of information as it arrived in the library and inform his client of anything of interest to him. This, of course, is the sort of personal service that many active information officers and librarians have been providing for many years for a small number of selected people in their organizations. Unfortunately, a personal service such as this can only be done for comparatively few people, and what is required is something on a much larger scale, for tens of thousands of persons instead of tens. Further, unless a great number of information officers were available to work on an organized basis, they could not hope to cover the vast flow of information.

3. Selective Dissemination of Information

A solution to the problem of providing a personal current awareness service to a large number of individuals was first suggested in 1959 by the late H. P. Luhn of I.B.M. in the United States. He called it the 'Selective Dissemination of Information', or S.D.I. for short.†

My subject interests are:

1. Mathematical theory of computers and automata
2. Pattern recognition and perception
   - Pattern matching
   - Human visual perception
   - Speech recognition
   - Cell recognition
   - Character recognition, etc.
3. **Learning systems**
   - Statistical learning theories
   - Association
   - Adaptation (adjustment of internal parameters)
   - Human and animal learning behaviour
4. **Problem solving**
5. Memory and information retrieval
6. **Cybernetics**
7. System reliability through redundancy
8. Systems using artificial neurons. Perceptrons

Fig. 1. User's statement of interests.

User's own words	*Words used in the S.D.I. system*
Learning systems	Learning
	Adaptive systems
	Self-organizing systems
Problem solving	Heuristics
Cybernetics	Cybernetics
	Bionics
	Man-machine systems

Fig. 2. Translation of user's words.

Luhn realized that one needed some means of comparing the subject contents of documents with the subject interests of the individuals. One could then send them details of the documents which matched their interests, that is those documents which seemed likely to be of interest to them. As this required an extremely large number of matching operations to be carried out very quickly, he developed a system for matching by computer.

† H. P. Luhn, IBM Advanced Systems Development Division Yorktown Heights, N.Y., 1959 (19 pp.).
Also in *American Documentation*, **12**, No. 2, pp. 131–8, 1961.

In the S.D.I. system each user supplies details of his subject interest in his own words. One such statement of subject interests we have received is shown in Fig. 1. The words used are then translated into the standard words used in the S.D.I. system. As an example of this, three of the user's words in Fig. 1 are translated in Fig. 2. Another example of subject-interest requirements and the standard words is shown in Figs. 3 and 4.

I am interested in:

1. All the literature on ZnS, ZnSe, CdS and CdSe
2. **Growth of crystal from the vapour phase**
3. Optical and electrical properties which give information on the **defect levels** in other semi-insulators (i.e. those compounds which have band gaps between about 1·5 eV and 3·0 eV): typical examples are absorption and emission spectra, thermoluminescence, photoconductivity, electrical conductivity as a function of temperature, space-charge-limited currents, measurement of drift mobility as a function of temperature, Hall effect, e.s.r.
4. Theoretical work on photoconductivity, double-injection electroluminescence, acousto-electric interactions
5. Solid-state photoconductive radiation detectors, image intensifiers, thin film triodes

Fig. 3. User's statement of interests.

User's own words	*Words used in the S.D.I. system*
Growth of crystal from the vapour phase	Crystal growth
	Vapour deposition
	Epitaxy
	Crystals
Defect levels	Crystal defects
	Vacancies
	Dislocations
	Interstitial atoms
	Frankel defects
	Traps
	Carrier-recombination centres

Fig. 4. Translation of user's words.

The standard words used in the S.D.I. system are incorporated in a controlled list of terms (or 'Thesaurus') in which synonyms have been removed, each word has a specific meaning, and the relationship between the words is clearly defined. An excerpt from the Thesaurus is shown in Fig. 5. Under 'Crystal defects' there is the instruction, conveyed by the abbreviation UF, that this is to be Used For 'Lattice defects'. The information that 'Defects' is a term of broader meaning is conveyed by BT (Broader Term), whereas 'Dislocations' and 'Vacancies' are shown to have narrower meanings, indicated by NT (Narrower Term). 'Crystal lattices' is shown to be a related concept by the letters RT (Related Term). Similarly under 'Lattice defects' there is an instruction to use 'Crystal defects' in its place.

Crystal defects

UF	Lattice defects
BT	Defects
NT	Dislocations
NT	Vacancies
RT	Crystal lattices

Lattice defects *use* Crystal defects

Fig. 5. Thesaurus.

The words defining a user's subject interests, together with his name and address, form his user-profile. Figure 6 gives the user-profile of a mythical Dr. Martin of R.A.E. who has interests ranging from

Dr. J. D. Martin

Radio Section, Royal Aircraft Establishment, Farnborough, Hants.

1 Digital circuits. 2 Electric circuits. 3 Time measurement. 4 Process control. 5 Synchronism. 6 Multivibrators. 7 Flip-flops. 8 Trigger circuits. 9 Logic circuits. 10 Counters. 11 Sequential circuits. 12 Remote control. 13. Radio telemetry. 14 Telemetry. 15 Thyristors. 16 Silicon controlled-rectifiers. 17 Electric switches. 18 Switching circuits. 19 Gate turn-off. 20 Diode switches. 21 Semiconductor devices. 22 Ultrasonic frequency. 23 Transducers. 24 Vibration. 25 Mechanical shock. 26 Sensors. 27 Magneto-striction. 28 Piezo-electricity. 29 Ferroelectricity. 30 Piezo-electric crystals. 31 Data storage. 32 Computer storage. 33 Electro-mechanical filters. 34 Acoustic amplifiers. 35 Pulse-code modulation. 36 Voice communication. 37 Audio-frequency. 38 Telephone systems. 39 Telephone transmitters. 40 Telephone receivers. 41 Coding. 42 Coders. 43 Decoding. 44 Decoders. 45 Compandors. 46 Analogue to digital convertors. 47 Digital to analogue converters. 48 Comparators. 49 Field effect transistors. 50 Transistor amplifiers. 51 Broad-band. 52 Measurement. 53 Distortion. 54 Inter-modulation. 55 Cross-modulation. 56 High frequency. 57 Microwaves. 58 Amplifiers. 59 Parametric amplifiers. 60 Masers.

Fig. 6. User-profile.

digital circuits to masers. The complete set of user-profiles are stored in a machine file, let us say on a reel of magnetic tape.

As the documents (which may be periodical articles, reports, conference papers or any other information material) are received by the information service they are indexed by subject. That is, one of the information workers examines the document, decides what subjects it is dealing with and adds the system-language words to the details of the document. In Fig. 7 the details given of the document are the title of the article, the names of the two authors, and an abbreviated form of the name of the periodical. Ten words or phrases have been added by the information workers to define the subject-matter of the document. These document details and the list of words defining its subject are then stored on a second reel of magnetic tape.

At suitable intervals the two tapes are compared in the computer and where sufficient of the words of a user-profile agree with the subject words associated with a document, the details of the document are printed out and sent to the user.

The system, however, does not stop there, since there is a feedback loop which allows the system to be continually adjusted to meet the users' requirements more accurately. Whenever the user receives a notification that a document is likely to be of interest to him, he is asked to say whether it is, in fact, relevant to his interests. The information he provides, his 'Relevance Assessment' as it is called, is used to modify his user-profile so that the system continually improves the service it provides to him. Similarly, if he requires a new category of information this can be added, or an unwanted category deleted.

(Title) **Investigation of some properties of a tunnel-diode amplifier.**

(Authors) V. P. Voronenko and A. N. Kovalev.

(Citation) Radio Engng. Electronic Phys. (U.S.A.), Vol. 10, No. 3, 385–391, March 1965.

(Subject words)	Tunnel diodes	Mathematical analysis
	Amplifiers	Current-voltage characteristics
	Saturation	Frequencies
	Polynomials	Load impedance
	Higher-order	
	Fifth-order	

Fig. 7. Document details with subject words.

An outline of the S.D.I. system is shown in the Flow Diagram in Fig. 8. The documents enter on the left, are indexed and incidentally photo-copied. The details of the document and the associated subject words form the Main Document File in the computer. Since all of the document details are not required for matching, the document numbers and subject words can be extracted from the Main Document File to form the Secondary Document File in the computer.

The subject interests of the Users, shown on the bottom right of the diagram, are translated into User-Profiles to form the User-Profile File which is then compared with the Secondary Document File. Where there is a satisfactory degree of matching between the items in the two files, the document number in the secondary document file is used to call up the complete details of the document from the Main Document

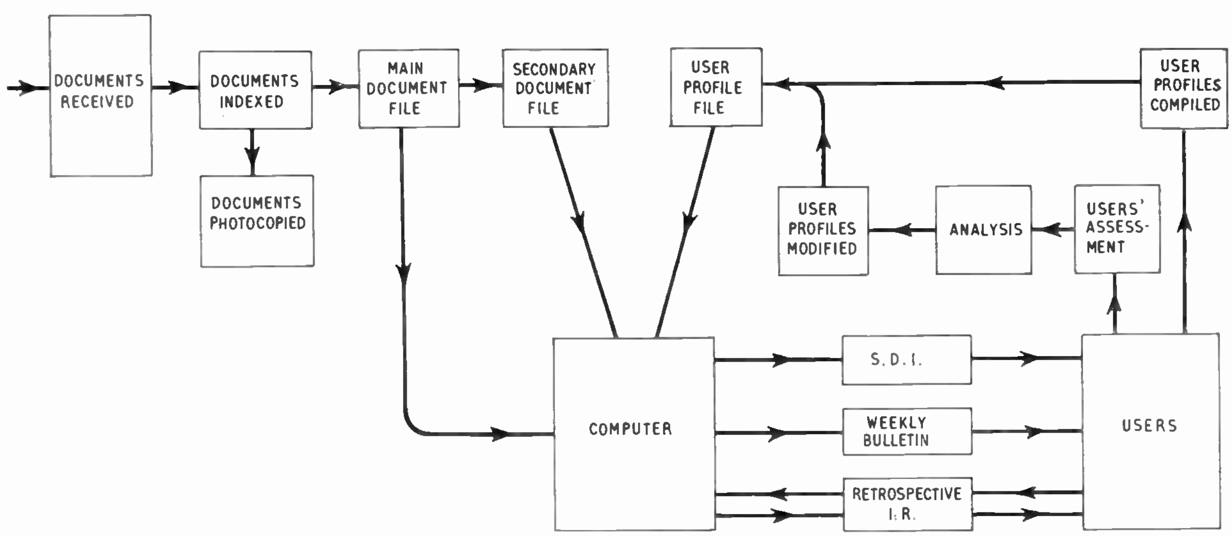


Fig. 8. Flow diagram for S.D.I. system.

File. Thus the S.D.I. service can provide full details of the documents as output. The other outputs provided are a weekly bulletin and retrospective searching, i.e. information retrieval.

The feedback loop is that linking the following boxes: Users, Users' Assessment, Analysis (that is, deciding what modification is required), User Profiles Modified, User Profile File, Computer and S.D.I.

4. The National Electronics Research Council

Before I discuss the National Electronics Research Council's S.D.I. Project, I might perhaps say a little about N.E.R.C. itself.

I have been very interested in electronics since I qualified as a naval signal and wireless specialist in 1924, and have been for many years a member of both the Institution of Electronic and Radio Engineers and the Institution of Electrical Engineers. In my Presidential Address to the I.E.R.E. in 1946 I indicated the value of research into computer techniques for information storage.† In another speech 15 years later I sowed the seed from which N.E.R.C. has grown, and it is interesting that N.E.R.C.'s first research project should be concerned with the use of computers for information work.

The speech in 1961 at which I put forward the idea of a national research organization in the field of electronics was made at a dinner given by the British radio industry.‡ I suggested to the industry that there was a need for co-operative research in the radio and electronics industry and subsequently asked the Institution of Electronic and Radio Engineers to conduct for me a survey on radio and electronics research in Great Britain.

I discussed the facts and figures shown in that survey with members of the Government of the day, the Vice-Chancellors' Committee of the British Universities, and the Chairmen of a number of British radio and electronic manufacturing organizations. We all agreed on the urgent need for establishing a central co-ordinating body which would concern itself with indicating to government and industry gaps in the British research effort in electronics, show where additional effort was required, recommend priorities, and suggest how to prevent unnecessary duplication of research.

Now it is not easy to bring together all the organizations which are concerned with electronics research in this country, and it is indicative of the importance of the National Electronics Research Council that it does, in fact, as shown in the Appendix, comprise representatives of all the government departments concerned with electronics research, including the government research stations, of the universities in Great Britain which are in any way concerned with electronics research, and also the Royal Society, and the other learned Institutions in this field, together with, at Chairman level, representatives of the British electronics manufacturing and aircraft industries. I make particular mention of the aircraft industry because without electronics we would not be able to utilize modern aircraft to full effect—in fact I doubt whether they could carry out a worthwhile flight.

N.E.R.C. has therefore succeeded in bringing together, for the first time in our history, all the leading figures who are able to determine the future course of electronics research in Great Britain.

From the first discussions on the possibility of forming N.E.R.C. we were convinced that one of the ways in which we could make a decisive contribution to the advancement of electronics research and perhaps also in other fields was by tackling the information problem. In fact, the Council established, as its first Working Party, the Working Party on Scientific Information, some four months before N.E.R.C. itself came into existence officially.

The Working Party lived up to its name. It immediately went to work to produce a survey of the state of the provision of information for electronic research workers during 1963. It estimated that some 25 000 documents of interest to electronics research workers were produced in that year throughout the world. Of these some 15 000 were either originally written in English or translated into English.

Its investigations showed that the facilities available for retrospective searching were very good for some types of material and, in general, were at least adequate. In particular the Working Party considered that the I.E.E. publications, *Physics Abstracts* and *Electrical Engineering Abstracts*, were excellent abstract journals and searching tools.

For current awareness on the other hand, the position was found to be thoroughly unsatisfactory. If a research worker wished to ensure that he was aware of the existence of all the periodical articles and reports of interest to him he might have to scan more than 10 abstract journals and title lists each month which might contain up to 3000 items in one issue. (The figure of 3000 has actually been taken from a recent issue of *Physics Abstracts* which includes a wider field than electronics.) Of course, in practice, no research worker would have time to do this even if he so wished.

The Working Party therefore considered that N.E.R.C. should concentrate initially on this current awareness problem and recommended an investigation of S.D.I. to deal with this problem.

† *J. Brit. I.R.E.*, **6**, No. 6, pp. 221–25, December 1946.

‡ *J. Brit. I.R.E.*, **23**, No. 1, pp. 3–6, January 1962.

5. S.D.I. Investigation

Although a small number of S.D.I. systems have been established, most of them in the U.S.A. within the last year or so, all of them have been limited either to the documents issued by the organizations which established them or to members of these organizations. It is our intention to include in the N.E.R.C. investigation all information material of interest from whatever source it may come, and to provide the service free during the investigation to research workers in government research laboratories, industrial firms, and universities and colleges of technology.

However, the main difference between the N.E.R.C. Project and the systems in the United States is that ours is specifically designed as an investigation. In other words, whereas the other systems were set up from the very beginning as operational systems, we shall be setting up an S.D.I. system for a select number of users only, to investigate the system with a view to improving and perfecting it.

The main purpose of the Project is to investigate the value, economics and acceptability to users of an S.D.I. system. All our reasoning leads us to believe that S.D.I. should be an extremely valuable technique for ensuring that each research worker in electronics is regularly informed of those items in the literature and only those items which are of interest to him. But no-one can tell how valuable the users will find such a service or whether it can be provided at an economic cost. The latter can only be judged in the final analysis on whether the cost of providing the service is less than the maximum amount the users are willing to pay.

Finally there is the question of acceptability to the user. This is extremely important as the S.D.I. service information is in a different form from the usual media. Information is normally issued in printed pages, in newspapers, periodicals, reports or books. Even if there are no advertisements or illustrations to brighten up the pages, many of the words in these publications are highly redundant and carry very little information content for any single individual. This is why one enjoys reading newspapers and why abstract journals, with the information boiled down to a few words, are so hard to read.

Similarly, we have to investigate what happens when a user is provided with a note saying, for example, that a paper entitled, 'Tunnel diodes integrated with microwave antenna systems', by H. H. Meinke has appeared in the February 1966 issue of *The Radio and Electronic Engineer*. If our system were working perfectly, each such notification would be of high interest to him, and there would be few unnecessary words: so it would require real concentration to read it. There would be no relaxation in looking at the pictures or reading through entertaining or superfluous words to find something useful. We don't know yet—and no-one else does either—whether this absence of redundancy will be acceptable or prove too great a discouragement. But we shall know at the end of the Project.

6. Details of the S.D.I. Project

(The next sentence is an example of one with low redundancy but even then it contains only 10 facts in 45 words.) The project is planned to take three years during which we shall set up an S.D.I. system to serve some 800 to 1000 research workers in electronics by providing them weekly with notifications of English-language periodical articles, conference papers and reports of interest to them.

6.1. *Users*

These research workers will be selected volunteers from industrial, governmental and academic research laboratories. We intend to make this group as representative as possible in the type of research they are engaged on and as varied as possible in their status, ranging from junior engineer to chief engineer and from research student to professor.

6.2. *Information Material*

Our reason for limiting the material used in the present Project to the English language was simply to avoid the extra effort and staff required to cover and translate foreign-language material. Apart from the straightforward translation difficulty, foreign-language material creates no special problem. Incidentally, our English-language material will include a large quantity of foreign material which is published in cover-to-cover translations under British or American government grants.

As I have said, the survey carried out by the N.E.R.C. Working Party showed that the total number of English-language items of interest to research workers in electronics is of the order of 15 000 a year at present. We intend to be as comprehensive as possible and to include all items of significance, but where possible to ignore popular or general technical articles which do not add to scientific and technical knowledge.

In referring to 'adding to scientific and technical knowledge', I should like to emphasize how important I consider it that the results of research work should be made freely available. Many British organizations engaged in electronics research are extremely liberal in their attitude to the publication of research results, mainly in the journals of the learned societies. British learned societies and similar bodies make no charge for the publication of papers giving the results of research work. In the United States, on the other hand, similar societies make 'page charges' for the inclusion of such material in their publications.

Organizations conducting research have an obligation, as members of that *élite* body, the scientific community, to play their part in disseminating information within the community. Only if the results of research work are freely published will the health and advancement of science and technology be assured and the standards of basic science maintained by critical review in the literature.

But self-interest should also play an important and natural part. It is widely recognized that it is in the organizations' own interests as well as in the national interest to publish their research results. Firms and other organizations with a reputation for encouraging the publication of results acquire a prestige which greatly increases their prospects of attracting the best type of research worker.

On the other hand, it is both unnecessary and impossible to keep research results secret. It is unnecessary because most of the important pieces of research are done elsewhere or superseded within a very short time. It is impossible since any research worker, in moving to a new organization, must inevitably take with him the knowledge of the work he has already done.

There are thus good reasons why organizations should make the results of their research freely available, but I believe more should be done by the government departments to provide incentives for this. In particular they should encourage the production and dissemination of scientific and technical reports on work done under contract. Government agencies in the United States insist on the production of such reports. For example the U.S. Department of Defense are stated to withhold 10% of the contractor's fee until the reporting requirements of the contract are fulfilled.

The U.S. government agencies are also extremely active in the dissemination of scientific and technical reports. Only last year they very substantially increased their already massive effort in this direction. They issue a number of abstract journals which list both the agencies' and their contractors' reports, and have established centres from which they can be obtained quickly and easily. A similar system for the announcement and distribution of scientific and technical reports in this country would be of great value and, I hope, will be developed. I suggest that it is essential to develop it as soon as possible and on an adequate scale.

In our own field of electronics research, we in N.E.R.C. are at present doing what we can to improve the exchange of information on research work. We have invited organizations concerned with electronics research to send us copies or details of their reports and publications for announcement in *N.E.R.C. Review*, our quarterly review of progress in electronics research. It is not our intention to set ourselves up as a lending library or distribution centre. Normally we merely list the reports and give the address of the organizations from which they may be obtained. Thus we do not wish *N.E.R.C. Review* to compete with the professional and other scientific

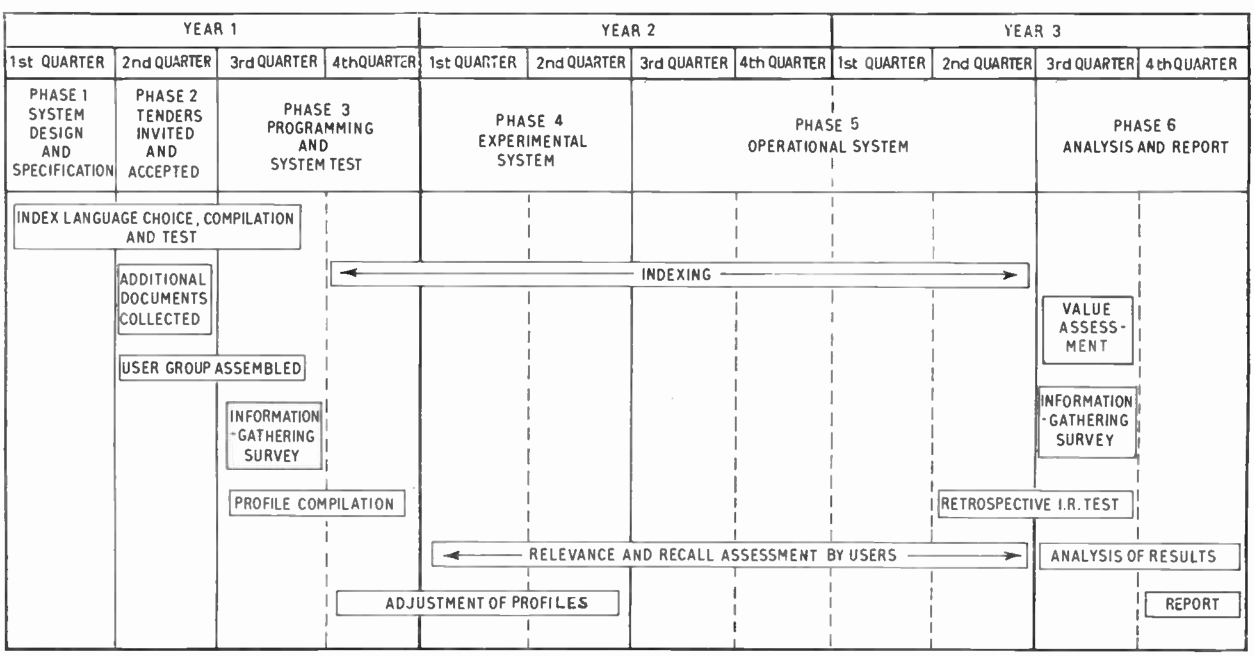


Fig. 9. Programme for the S.D.I. project.

journals which give detailed reports of new research work. What we seek is early advice of research achievement so that we can distribute such information quickly; the more detailed reports can appear later in the scientific and technical journals. We shall, of course, include all of these reports in our S.D.I. Project.

6.3. *Computer Aspects*

We shall put out a contract to a computer bureau who will design programs to meet our specification, do the programming, and carry out the weekly runs.

At this stage it might be interesting to have a look at our programme of work (Fig. 9). This was drawn up almost 18 months ago and I am glad to say is still sound enough for us to work to.

The first row of boxes under the heading 'Quarters' show that we have six months to complete Phases 1 and 2 and get to the stage where the computer bureau can start the detailed programming and system test of Phase 3 which takes another six months. That completed, we start the system and run it for six months on an experimental basis as Phase 4. Of course, we are ready to meet a lot of trouble and to make extensive changes to the user-profiles during this period.

Then when, as we hope, the system has settled down, there will be a year of proper operational working during Phase 5. It is during this time that we shall carry out most of the tests on user reaction to find out how well we are doing the job. Finally we have allowed ourselves six months in Phase 6 for the analysis, and to compile the final report.

6.4. *S.D.I. Notifications*

Each week, the users will receive notifications of items of interest to them, but not the original items or photocopies. The notifications will contain information on the lines shown in Fig. 10. This contains details of the same document as in Fig. 7, but in the form in which it would be sent out. It gives our

---

003019

Investigation of some properties of a tunnel-diode amplifier.

V. P. Voronenko and A. N. Kovalev.

Radio Engng. Electronic Phys. (U.S.A.), Vol. 10, No. 3, 385–391 March 1965.

Tunnel diodes. Amplifiers. Saturation. Polynomials. Higher-order. Fifth-order. Mathematical analysis. Current-voltage characteristics. Frequencies. Load impedance.

Fig. 10. Notification of document.

74325

Mr. J. D. Carlsfield.

Central Research Laboratory, Wilson Electronics Co. Ltd., Causeway Avenue, Harpenden, Herts.

Via: Mr. R. C. Robertson, Librarian.

003019 003065 003187 003279

Fig. 11. S.D.I. output (1).

74325

Mr. J. D. Carlsfield.

Central Research Laboratory, Wilson Electronics Co. Ltd., Causeway Avenue, Harpenden, Herts.

Via: Mr. R. C. Robertson, Librarian.

003019

Investigation of some properties of a tunnel-diode amplifier.

V. P. Voronenko and A. N. Kovalev.

Radio Engng. Electronic Phys. (U.S.A.), Vol. 10, No. 3, 385–391, March 1965.

Tunnel diodes. Amplifiers, Saturation. Polynomials. Higher-order. Fifth-order. Mathematical analysis. Current-voltage characteristics. Frequencies. Load impedance.

003065

Thin-film preparation.

D. Dobischek and S. Cabell.

Army Electronics Lab. Report AD–614779.

Thin films. Vacuum deposition. Tunnel diodes. Aluminium. Magnesium oxide. Tunnelling (semiconductors). Emitters (electron).

003187

Development of experimental electronically-tuned, decade-bandwidth receiver.

B. Barber and others.

Melabs, Palo Alto, Calif. Report AD–614971.

I.F. amplifiers. Radio receivers. Radar receivers. Electronically tuned. Tunnel diodes. Solid-state devices. Broadband. Ultra-high frequency.

003279

An improved construction for tunnel diodes.

M. J. Coupland and others.

S.E.R.L. Technical Journal, Vol. 12, No. 2, 98–104, June 1962.

Tunnel diodes. Fabrication. Gallium arsenide. Diode switches.

Fig. 12. S.D.I. output (2).

code number for the document, the title and authors, the citation (i.e. in which journal it appeared or in which report series), and a list of the terms we used to define the subject of the document, instead of an abstract.

At present we do not know whether we shall send each user a printed-out list of document numbers only (Fig. 11). With the list we would send him a bundle of cards, one for each document. The alternative is to send him a print-out of the full details of his week's notifications on a sheet or sheets as shown in Fig. 12. There are good reasons for both types of notification; we shall not decide which to use until we know exactly the relative costs and value of each method.

Whichever method we use, we shall also send for each item notified a pre-scored card (Fig. 13). The user can punch holes at the pre-scored positions with his pencil. He will be asked to return it as soon as possible to tell us whether the item is relevant to his interests or not. This relevance assessment is the feedback which is used to improve and finally perfect the information received by each user. In our Project it also plays an important part in carrying out our investigation.

Either

- I HAD ALREADY SEEN THE DOCUMENT ☐
- It is relevant to my interests ☐
- It is not relevant to my interests ☐

Or

- MY ASSESSMENT IS BASED ON THE DETAILS PROVIDED ☐
- The item is relevant to my interests ☐
  - and I intend to examine the document ☐
- The item is not relevant to my interests ☐

Or

- MY ASSESSMENT IS BASED ON EXAMINATION OF THE DOCUMENT ☐
- The document is relevant to my interests ☐
  - and is useful to me ☐
  - but it is not useful to me for the following reason ☐

  ....
- The document is not relevant to my interests ☐

Fig. 13. Relevance assessment.

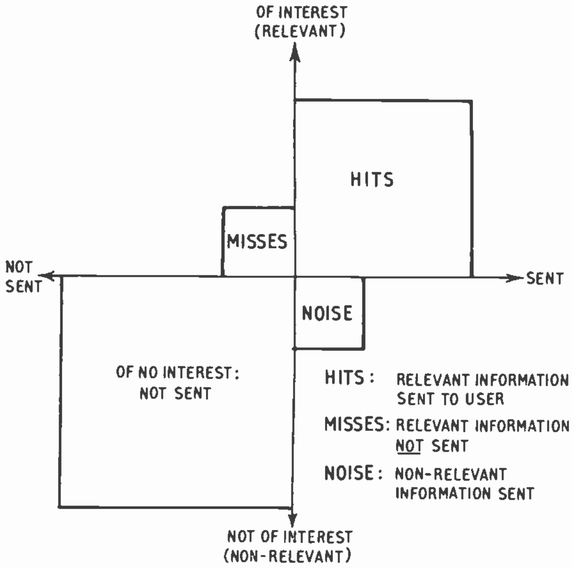


Fig. 14. Test for faults arising in S.D.I. system.

6.5. *Analyses and Tests*

The analyses and tests are, of course, the essence of our Project. The first we shall measure is the general performance of the system. We want each user to receive all the information of interest to him but no other material. So from his relevance-assessment cards we can find what percentage of the notifications we sent him were not relevant to his interests.

But, of course, this is only half the problem and the easier half for which to obtain an answer. We also want to know what we should have sent him but for some reason failed to do so. We shall obtain this answer by asking each user to scan, say every month, a copy of our weekly bulletin and let us know what he feels he has missed. The weekly bulletin will be a complete listing, in a broad subject order, of all the information material that was received by the system during the week.

From these two returns we shall be able to work out the relative areas of the boxes in Fig. 14. The diagram has no pretensions to scientific merit, and there is no scale. It is simply an attempt to show the two types of fault that arise in the system. The area above the horizontal axis contains information which is of interest to the users, that is, relevant information: the area below it contains information which is not of interest, or non-relevant. On the right of the vertical axis is the information that has been sent: all that has not been sent is on the left.

In the top right we have the HITS, the information which the users wanted to see and which was sent to them. In the bottom left is all the very large mass of information that was of no interest and was not sent.

The other two squares represent the errors. At the top left, MISSES, information which the users would have liked to see but did not receive, and lastly, NOISE, the information they did not want to see but were sent. It is these MISSES and NOISE squares that we must make as small as possible.

The diagram is somewhat misleading, however, in that the horizontal axis is represented as a line. In practice this would be a band, of somewhat uncertain width, since the change from relevant to non-relevant is usually gradual and uncertain. This, of course, adds to the difficulties in assessing the performance of the system.

The machine file of information which is processed each week for the S.D.I. service will be stored and can be searched at any time by the computer: that is, a retrospective search can be made. Towards the end of the Project we shall be testing the efficiency of the system for such retrospective searching—there is little point in doing this until a reasonable quantity of information has been stored.

There is a large number of other analyses and tests we intend to carry out but I want to mention just one more, because of its general importance.

One of the serious problems in dealing with scientific and technical information is the apparent reluctance of scientists and engineers to make use of information services. I am told that in general they will ask anyone, telephone to the other side of the country, in fact do anything rather than use the library or information service provided by their organization or community.

We hope and believe that S.D.I., since it is a new and different technique, will alter their approach to information. To test this we shall carry out a survey of our S.D.I. users before we 'expose' them to S.D.I. and repeat it when they have been receiving the service for a year. To make sure that any difference is not due to a general change, unrelated to S.D.I., we shall similarly survey, on both occasions, a matched control group who will not receive the S.D.I. service.

7. S.D.I. in Other Countries

We have naturally been interested in finding out if any organization in this or any other country is conducting or planning an investigation of S.D.I. on similar lines to our own. So far as we know we have no competitors, and we have confirmed this with the government organizations in this country and in the United States which keep in touch with information research throughout the world. Of course, we shall continue to ensure that our current awareness of this is as complete and up-to-date as possible.

We are keeping in close touch with what the Americans are doing in S.D.I. They are well ahead of us in practical work but we believe that from our investigation we shall discover the design of system which best meets the users' requirements.

The Japanese are also interested in S.D.I. The President of the Japan Information Centre for Science and Technology, Mr. S. Hamada, is to come to London in February, and will spend a day discussing the whole question with us at the N.E.R.C. Headquarters.

Then there is the Russian work, about which we have so far failed to obtain any information. On the 13th January I had the chance of talking to Mr. A. N. Kosygin, the Russian Prime Minister, as we were both in New Delhi to attend the funeral of Lal Bahadur Shastri. When I told him about this lecture he expressed great interest and mentioned that the Russians had made great advances in dealing with scientific and technical information. I offered to send him a copy of my lecture and said that I hoped he would send me, in return, full particulars of the progress his people had made in this field. He promised that, if I wrote to him, he would give the matter his personal attention.

8. The Pilot Phase of the S.D.I. Project

A year ago last October, we submitted a proposal for the three-year project to the then Department for Scientific and Industrial Research. D.S.I.R.'s role in the support of research in information was taken over by the Office for Scientific and Technical Information of the Department of Science and Education (O.S.T.I.) and in May 1965 we received a letter of grant from them awarding us up to £6000 for support of the first five months of the Project. If we look back at the programme of work (Fig. 9) this five-month period is approximately Phases 1 and 2 of our Project.

We started work on the five-month pilot phase project on 1st October 1965. The aim of this phase is to complete the detailed specification of the S.D.I. system and to obtain tenders from computer bureaux for the design, programming and weekly operation of the system. We need this information to enable us to obtain definitive figures for computer costs for incorporation in our proposal for the remainder of the Project. We issued the specification to computer bureaux on 31st December and asked for tenders to be submitted by 11th February 1966.

This, however is only one aspect—the computer aspect. The S.D.I. Project will not stand or fall by the computer operation: essentially all the problems are the human, intellectual ones that are inherent in any information system however highly mechanized it may be.

We have therefore been investigating manually the system language, the compilation of user-profiles and the matching of these with the documents. For this investigation we have indexed by subject over 800 documents, consisting of both periodical articles and research reports, covering the broad spectrum of electronics research.

8.1. *Subject-profiles*

For subject-profiles we invited the co-operation of a number of librarians and information officers. We asked them if they would obtain details for us of the subject interests of a number of users of their information services and act as our agents, discussing their problems and answering their queries where they could.

We had hoped to obtain perhaps 40 specimen subject-profiles in this way. In fact, we have had very full co-operation from the librarians and information officers we approached. Similarly, their users have shown great interest in the Project so that we obtained details of the subject interests of more than 120 users in government research laboratories, research associations, universities, colleges of technology, and industry.

As shown in Fig. 15, we translated their subject interests, given in Column 1, into our system-language

Column 1 Terms of description of interests supplied by user	*Column 2* Translated into indexing language	*Column 3* Suggested additional terms (delete those not applicable)
1. Radar systems	Radar	Pulse doppler radar
	Radar equipment	Doppler radar
		Monopulse radar
		Pulse radar
		Radar transmitters
		Radar receivers
		Radar antennae
2. Systems employing radio, optical or sonic propagation	Lasers	Sonar
	Sound waves	Ultrasonic frequency
	Sound ranging	
	Sounding	Echo sounding
3. Systems designed for land, sea or airborne use		Ground controlled approach
		Radar navigation
		Air navigation
		Doppler navigation
		Marine navigation
4. Systems designed to detect, locate, track or measure the velocity of moving or fixed targets	Detection	
	Radar scanning	
	Radar tracking	
	Velocity measurement	
	Range finders	
	Distance measurement	

Fig. 15. Compilation of user-profile.

(Column 2), and returned the details to the users with suggestions of associated terms and possible additional concepts (Column 3). When the user-profile had been agreed by each user, it was compared manually with the 800 indexed documents. Then copies of these documents which gave a simple match were sent to the user. From his assessment of the relevance of the documents we were able to analyse the efficiency of different methods of matching.

8.2. *Matching*

This question of matching is naturally extremely important, since it is the basis of S.D.I. There are a number of different ways in which matching can be carried out, but I think I can most easily show the wide variation by describing the simplest and the most sophisticated.

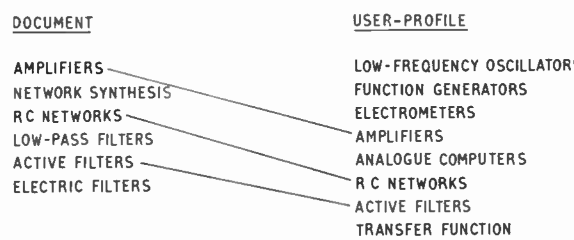


Fig. 16. Matching on three terms.

The simplest method is where one, two or more of the words must be identical for the match to be successful. In Fig. 16 you will see that three words are identical. If that were the condition of match specified, this would be a successful match.

The more sophisticated method used Boolean logic in the manner illustrated in Fig. 17. A simple example

$1+2.(3+4+5.6).\bar{7}$

That is, an acceptable match will result if either:

Term 1 appears, even on its own, or
Term 2 appears with either Term 3 or Term 4 or with Terms 5 and 6 together, so long as Term 7 is not included in any of these combinations.

Note: In Boolean logic the following signs are used:

$+$ represents 'or'
. represents 'and'
$\bar{1}$ (a bar above a figure) represents 'not' that figure
( ) have the normal algebraic use

Fig. 17. Logical statement.

of such a Boolean logical statement is shown in Fig. 18.

From this it will be appreciated that the operation of an efficient S.D.I. system is more involved and beset with more difficulties than would appear from the simple principle on which it is based. On the other hand, as we proceed we continue to find it an even more fascinating problem than we ever imagined.

1. Distortion
2. Microwaves
3. Amplifiers
4. Masers

1.(2.3+4)

i.e. distortion and ((microwaves and amplifiers) or masers) or, as normally expressed, 'distortion in microwave amplifiers or masers.'

Fig. 18. Use of logical statement.

At the end of February, therefore, we hope to have an interesting and encouraging report to make to N.E.R.C. and to O.S.T.I. on the results of our five months' work. What is more important to us is that we hope to be in a position to submit a proposal, in considerable detail, of the programme of work we wish to do next. So at the beginning of March we hope to have the go-ahead to embark on the main part of the Project.

9. Wider Implications of S.D.I.

So far I have been discussing S.D.I. in the context of what we are doing on our present pilot-phase project and what we intend to do in the main part of the Project in the following two to three years.

But although we must devote all our efforts to these immediate goals we have continually in mind the great potentialities of the S.D.I. system and its use on a much larger scale in a wider context. It is with these possible developments of the system, some of them still years away, others almost complete, that I should like to deal in the final part of my discourse.

Our present project, since it was initiated by N.E.R.C., is tackling the information problem in the field of electronics research. Electronics research is a very suitable field in which to conduct the investigation as we have close relations with all the organizations concerned with electronics research and know that we can obtain the co-operation of their research workers. But there is nothing in the S.D.I. system that requires it to be confined to information on electronics.

Similarly, there is no reason why the system should not spread far beyond the confines of this country, in particular throughout the Commonwealth.

9.1. *Commonwealth Participation*

The whole concept of the National Electronics Research Council has aroused interest in the more advanced Commonwealth countries and a number of them are investigating the setting up of a similar body in their own countries. The S.D.I. Project has a particular appeal to these countries since they realize that the information problem cannot be tackled satisfactorily or easily by any one of them on its own.

Last year, while still serving as Chief of the Defence Staff, I had occasion to visit Canada, Australia, New Zealand and India. Our literature and all recent numbers of *N.E.R.C. Review* had already been sent to the appropriate people but I was agreeably surprised to find the intensity of interest shown during my visit.

I welcomed the inquiries made by government departments, universities and industrial organizations. My discussions with them have led us to examine how N.E.R.C. can bring in Commonwealth participation in the S.D.I. Project.

Some months ago, the Weapons Research Establishment in South Australia offered to nominate several of its research workers to take part in our investigation. Then, as recently as last November, the Secretary of the Institution of Electronic and Radio Engineers, Mr. Graham Clifford, who was one of the founder members of N.E.R.C., visited Canada and had valuable discussions on behalf of N.E.R.C. As a result he has been able to arrange for some 70 to 80 Canadian research workers in electronics to take part in the later phases of the S.D.I. Investigation.

Mr. Clifford is shortly visiting India, Australia and New Zealand and will discuss on our behalf how research workers in those countries can take part in the later phases of the Project.

Thus we hope to have the help of up to 200 research workers in other Commonwealth countries. These will be additional to the 800 United Kingdom participants in the main Project, bringing the total to 1000. We hope that their participation will allow us and them to assess the importance of their special problems of distance and time, and perhaps availability of the actual information material. It will certainly give us a wonderful opportunity to forge close links in co-operative activities in information work.

This could be particularly useful since we see a two-way flow of information. From the United Kingdom there would be the flow of S.D.I. notifications to the research workers in the different countries. From these countries we hope to be supplied, for inclusion in the

S.D.I. system, with unpublished or semi-published material issued by their various government establishments, universities and industrial organizations. These will mainly be research and technical reports, university theses, and conference papers. Few of these are made generally known or available in this country, or indeed other countries. We hope that each country will nominate a liaison officer who will acquire the documents on our behalf.

We feel strongly in N.E.R.C. that this bonding together of Commonwealth research interests will not merely be valuable: it is really essential if we are to compete with the vast resources of the U.S.A. and U.S.S.R.

9.2. *Application to other Subjects*

As mentioned earlier, there is no reason why the S.D.I. system should be confined to the subject area of our investigation, electronics. In fact, our investigation is designed with the express intention that, if it is successful, the system developed for the Project (that is, the concepts, methods and computer programs) and the results of the investigation, will be immediately applicable to any other branch of science or technology

In the field of medical information, computers are in regular use in the United States for the production. of the abstract journal, *Index Medicus*, and for information retrieval. A British system is being established at present to use the computer tapes generated by the U.S. system. These tapes contain both the details of documents and the words defining the subject matter of the documents. Thus one half of the input required for an S.D.I. system is already available. Only a file of user-profiles would have to be added to the system to establish an S.D.I. service.

In electrical engineering and physics, the fields covered by the Institution of Electrical Engineers' publications *Electrical Engineering Abstracts* and *Physics Abstracts*, an S.D.I. system could usefully supplement the computer system which is projected for producing their abstract journals and providing ancillary information services.

In chemistry and chemical engineering, with their especially large and growing literature, S.D.I. would be particularly useful. There is the problem of codification of chemical structures but much work has been done on this and a number of solutions are available.

There is similarly scope for S.D.I. in other sciences such as biology, oceanography, mathematics and statistics, but it is in the technologies that S.D.I. might provide a real breakthrough in information work.

Mechanical engineering and structures, aircraft engineering, civil engineering, metallurgy, transport engineering, production engineering and similar engineering subjects or technologies are extremely difficult subjects in their information requirements. There are few abstract journals covering each of these fields entirely and in general it is difficult to define the fields accurately and to ensure that all the information is covered. An S.D.I. system which covered the whole field of technology and engineering would overcome many of the present difficulties and be of very great help to engineers and technologists who at present are less well catered for and have much more of an information problem than scientists and research workers.

9.3. *Possible Later Developments*

The two developments I have mentioned—the participation of research workers in Commonwealth countries in our Project, and the application of S.D.I. to other disciplines—may occur within the next few years. But it is when one thinks of the possible development of S.D.I. further in the future, say 10 to 15 years ahead, that one can see most clearly its great possibilities.

Let us take the Commonwealth interests first. Rather than send copies of the notifications from this country to the research workers in the various countries, the system would be decentralized for speed and efficiency. Each country would have its own list of user-profiles on a computer file but all the subject-indexing of documents and computer programming would be done in the United Kingdom. Each week a copy of the tape containing details of the new documents would be transmitted to each country which would then provide an S.D.I. service to its research workers using its own computer.

In Britain, as data transmission techniques develop and research establishments acquire their own computers, one can visualize a more advanced but already feasible scheme of S.D.I. All scientific and technical information, whether in periodical reports, conference papers or any other form, would be subject-indexed at a large information centre. Each evening this subject-indexed material would be matched with the broad subject requirements of each organization taking part in the scheme.

Details of the documents which matched each organization's requirements would be transmitted direct from the centre's computer over a data transmission link to the organization's computer. In the latter a further matching would be made automatically with the profiles of individual scientists and engineers and the S.D.I. output distributed. Thus, every morning, each scientist or engineer would have waiting for him on his desk details of all the information of interest to him which had been received by the information centre in the previous day.

10. Conclusion

This lecture has not only been a wonderful experience for me but it has been of the greatest value to N.E.R.C. to have this opportunity of making known the details of the S.D.I. Project before the conclusion of the pilot phase. I believe that the fullest benefit cannot be obtained from our Project without an adequate knowledge of what we are trying to do, so that all who have the interests at heart of the future development of research in electronics and other sciences and technologies throughout the Commonwealth can give us at least their moral support. I propose to give a talk on S.D.I. in Canada in April and in India in a year's time to further this end.

11. Acknowledgments

I should like to pay tribute to the continual encouragement and help which N.E.R.C. has received from the staff of the Office for Scientific and Technical Information and from the Director, Dr. H. T. Hookway. I particularly wish to acknowledge the personal interest the Secretary of State for Education and Science, Mr. Anthony Crosland, and his Permanent Under-Secretary of State, Sir Herbert Andrew, have shown in the Project.

May I thank Lord Fleck, Sir Lawrence Bragg and all who have made this evening such a memorable one for me in this historic lecture theatre of the Royal Institution.

12. Appendix

Representation on General Committee of N.E.R.C.

MINISTRY OF AVIATION, by Permanent Secretary.

BOARD OF TRADE, by Permanent Secretary.

MINISTRY OF DEFENCE, by Permanent Under-Secretary of State and Chief Scientific Adviser.

DEPARTMENT OF EDUCATION AND SCIENCE, by
Chairman, Science Research Council, and
Principal, Northampton College of Technology.

POST OFFICE, by Director-General.

COMMITTEE OF VICE-CHANCELLORS AND PRINCIPALS OF THE UNIVERSITIES OF THE UNITED KINGDOM, by
Professor of Telecommunications, Imperial College, London.
Professor of Electrical Engineering, University of Glasgow.
Professor of Electronic and Electrical Engineering, University of Birmingham.

NATIONAL RESEARCH DEVELOPMENT CORPORATION, by Managing Director.

ROYAL SOCIETY

INSTITUTION OF ELECTRONIC AND RADIO ENGINEERS

INSTITUTION OF ELECTRICAL ENGINEERS

CONFERENCE OF THE ELECTRONICS INDUSTRY, by
Chairman, Elliott-Automation Limited.
Chairman, Plessey-UK Limited.
Chairman, Mullard Limited.
Chairman, Ferranti Limited.
Vice-Chairman, The General Electric Company Limited.
Chairman, Electric and Musical Industries Limited.
Director of Electronics Group, Associated Electrical Industries Limited.
Chairman, The English Electric Company Limited.
Chairman, English Electric-Leo-Marconi Limited.
Chairman, Thorn Electrical Industries Limited.
Director, Standard Telephones and Cables Limited.

SOCIETY OF BRITISH AEROSPACE COMPANIES LIMITED, by
Chairman, Hawker Siddeley Group Limited.

Manuscript received by the I.E.R.E. on 31st January 1966. (Address No. 36).

This Address is being published in the *Proceedings of the Royal Institution*

Meeting of the Institution's Canadian Division

On 14th April Lord Mountbatten addressed a meeting of the Institution's Canadian Division in Ottawa on 'The Selective Dissemination of Information Project of the British National Electronics Research Council'. His paper was based on the foregoing address, and the meeting was held in co-operation with the Canadian National Research Council and the Defence Research Board. The chair was taken by Professor A. D. Booth, Chairman of the Canadian Division. The Division gave a dinner in honour of Lord Mountbatten at the Chateau Laurier Hotel, the same evening.

U.D.C. 621.373.51

Design of Low-noise Solid-state Microwave Sources

By

J. FRILLEY†

AND

G. GRANDCHAMP†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July, 1965.

Summary: In applications such as communication and Doppler systems, the f.m. noise performance of microwave sources is quite critical. The noise performance of X-band microwave sources using transistorized crystal oscillators and amplifiers and varactor frequency multipliers has been studied. The paper discusses the limitations due to the crystal, the associated oscillator, the amplifier and the multiplier and also the methods used to measure their contribution. Theoretical and experimental results are compared.

1. Introduction

The subject of this paper is the analysis of the noise characteristics of typical solid-state microwave sources using a transistor v.h.f. oscillator and amplifier followed by a varactor multiplier.

The driver and local oscillators of high capacity communication systems have to meet very tight specifications regarding the f.m. noise accompanying the carrier. For Doppler radars, system engineers are usually interested in short term frequency stability or in short term overall phase jitter which can be calculated as soon as the f.m. noise spectrum and the demodulation transfer function of the equipment are known.

The contribution of each part of the chain to the overall parasitic angular modulation will be studied theoretically and experimentally in this paper. The theoretical limitations due to the crystal, the associated oscillator, the amplifier, the harmonic generator will be described as well as the experimental methods used to measure their contribution.

2. Definitions relative to F.M. Noise

The authors use for noise quantities the definitions most commonly employed by tube manufacturers. F.m. noise power is measured in a 1 kc/s band at a certain distance from carrier frequency. Then f.m. noise is expressed in terms of carrier frequency deviation in cycles/second (root mean square value).

The following example will clarify this definition:

Assuming that a power source has, with this definition, an f.m. noise of 10 c/s r.m.s. referred to a 1-kc/s bandwidth 100 kc/s away from the carrier frequency, then, in the reference bandwidth, the noise power is the same as the power which would be measured in one sideband if a noise-free source of the same carrier level was frequency modulated at 100 kc/s with a carrier frequency deviation of 10 c/s r.m.s.

† Thomson-Varian, 6 rue Mario Nikis, Paris, 15e.

Noise power is actually measured at video frequencies and noise figures are then referred to one sideband at radio frequency.

Then, with this definition, the quantity actually measured is the carrier/noise-power ratio which can be written to a first approximation as

$$\left(\frac{C}{N}\right)_{\text{dB}} = 20 \log \frac{\sqrt{2}F_m}{\Delta f}$$

Δf being the r.m.s. carrier frequency deviation, $F_m = f_1 - f_0$ with f_0 = carrier frequency, f_1 centre frequency of the 1 kc/s window; C is carrier power and N noise power.

This approximation is justified if the modulation index is small, an assumption which is always true in the case of modulation by noise. The expression takes into account the $1/\sqrt{2}$ ratio between r.m.s. and peak deviation. In Appendix 1 are given the relations between these quantities and those which are commonly used by system engineers, that is to say:

Noise power measured in picowatts referring to a 1 mW signal in a psophometric window of 4 kc/s bandwidth.

Signal/noise ratio expressed in decibels by comparison with a 200 kc/s carrier deviation for a 1 mW level.

Short term frequency stability.

Short term phase jitter.

3. Measuring Equipment

Most of the measurements have been performed on solid-state sources including crystal oscillators in the 100 Mc/s range and delivering 150 mW in X-band after a frequency multiplication by a factor 96 ($4 \times 3 \times 2 \times 2 \times 2$) (Fig. 1).

Most of the test equipment works in X-band and is housed in a screened room. It comprises:

(1) Microwave frequency discriminators which are associated with various wave analysers and a very

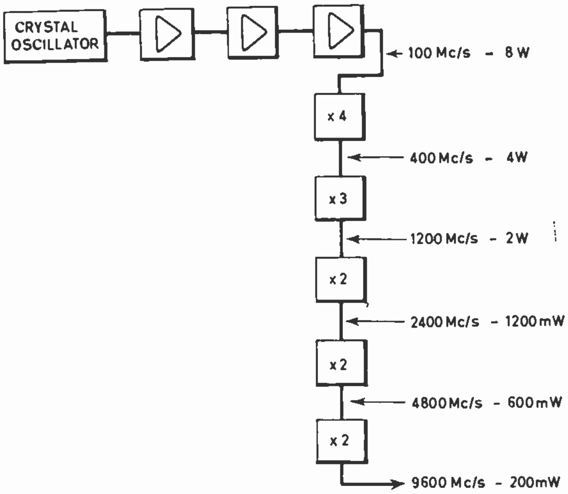


Fig. 1. Typical solid-state source used in the experiments.

sophisticated calibration apparatus. A typical discriminator is shown in Fig. 2. These discriminators are used to plot the f.m. noise spectrum. The typical sensitivity allows the measurement of noise frequency deviations as low as 0·1 c/s related to a 1 kc/s bandwidth.

(2) A differential measurement set-up which allows the measurement of the phase jitter between two sources having the same output frequency. The block diagram of this set-up is shown in Fig. 3. The typical sensitivity of this equipment permits the detection of phase components in a 1 kc/s bandwidth corresponding to a jitter of 3 seconds of arc(10^{-5} radian). The contribution of each part of the chain can be measured by the following process:

Identical solid-state microwave sources feed the two input waveguides of the bridge. Then, successively, various parts of the chains are driven by a common generator, the lower frequency parts of the chain left out of operation being disconnected. For each successive experiment the remaining phase noise is the contribution of the sections of the chains under test. By successive differences the noise contribution of the oscillator, the low level amplifier, the power amplifier and the varactor multiplier can be evaluated. Assuming that the two chains are identical, the result is decreased by a factor of 3 dB. If one chain is worse than the other the maximum error is 3 dB and this is quite acceptable in measurements of this type. The differential phase bridge includes an auxiliary phase-locking channel which is used to maintain long-term phase difference to a fixed value in certain experiments. The bridge is calibrated statically by feeding the same signal directly in one arm and through a precision phase shifter in the other arm.

4. Basic Experimental Results

The following results have been used as a guide for theoretical analysis of the noise spectrum.

Figure 4(a) shows the typical f.m. noise spectrum of a klystron. The deviation Δf related to a 1 kc/s bandwidth decreases as F_m increases (F_m being the difference between the centre frequency of the 1 kc/s window and the carrier frequency).

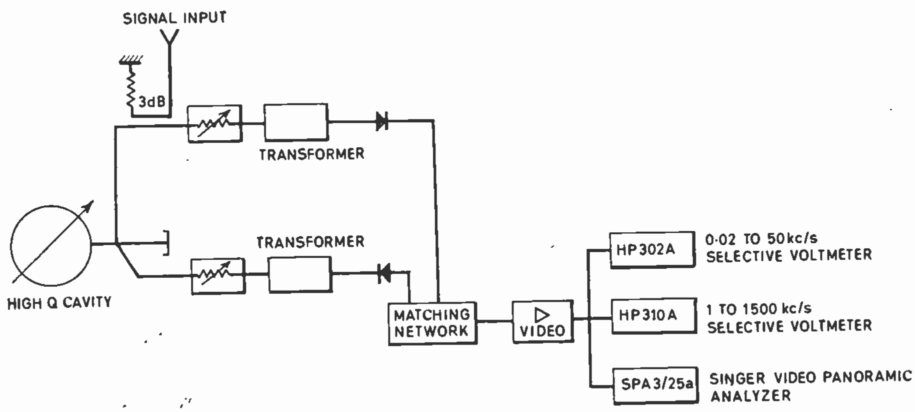
Figure 4(b) shows the typical f.m. noise spectrum of a solid-state varactor source which is quite different from that of the klystron.

Obviously the nature of the noise is not the same in the two cases. In the first case, noise sources in the klystron lead to pure f.m. noise which decreases continuously on each side of the carrier.

In the second case, the noise spectrum has been understood as being the sum of two contributions. The f.m. components which appear very close from carrier are due to frequency modulation in the oscillator.

This frequency modulation is related to the loop gain phase jitter. The importance of this contribution depends upon the crystal loaded Q. Replacing the crystal oscillator by a free running oscillator leads to a much wider spectrum as shown in Fig. 4(c).

Fig. 2. Typical microwave discriminator and video equipment.



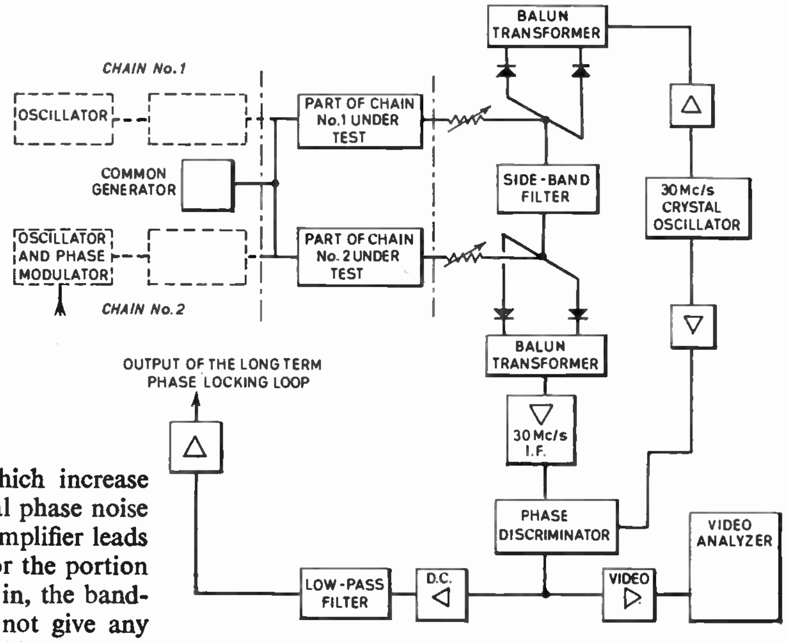
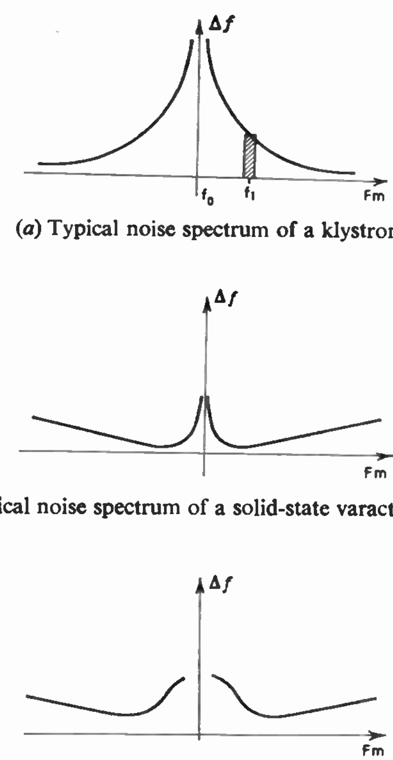


Fig. 3. Differential phase-jitter measurement set-up.

The components of the spectrum which increase linearly with F_m are due to the additional phase noise of the amplifier. Thermal noise in the amplifier leads to amplitude and phase modulation. For the portion of the spectrum which we are interested in, the bandwidths of amplifiers and multipliers do not give any limitation. The phase jitter is then a white noise of spectral density $\delta\varphi$, constant with frequency, and leads to an f.m. noise, the density δf of which is $F_m \delta\varphi$ and increases with F_m. The carrier/noise power ratio

$$\left(\frac{C}{N}\right)_{\text{dB}} = 20 \log \frac{\sqrt{2} F_m}{\Delta f}$$

is constant as F_m varies.



(*a*) Typical noise spectrum of a klystron.

(*b*) Typical noise spectrum of a solid-state varactor source.

(*c*) Typical noise spectrum of a solid-state source driven by a free running oscillator.

Fig. 4.

5. Theoretical Results

5.1. *Contribution of the Additional Noise of the Amplifier*

The assumptions are:

The amplifier is supposed to work in linear operation; that is to say the signal and noise functions can be handled separately and added at the output.

The varactor multiplier does not itself contribute to noise and its role is limited to multiplying the index of the parasitic modulation.

The r.m.s. frequency deviation is expressed by:

$$\Delta f = n F_m \frac{\sqrt{N}}{\sqrt{2\left(\frac{C}{N}\right)_{\text{in}}}} \frac{\sqrt{A(\omega)}}{\sqrt{A(\omega_0)}}$$

where

Δf = r.m.s. frequency deviation due to noise measured in a 1 kc/s bandwidth at F_m c/s away from the carrier.

N = noise figure of the amplifier

n = frequency multiplication factor

$A(\omega)$ = power gain of the amplifier

$A(\omega_0)$ = power gain of the amplifier at carrier angular frequency

$\left(\frac{C}{N}\right)_{in}$ = ratio of carrier-power to amplitude-noise-power in a 1-kc/s bandwidth at F_m kc/s from the carrier at the input of the amplifier. The use of this ratio takes into account the fact that thermal noise can be represented by two equal components in quadrature. The one which is in phase with the signal gives amplitude modulation. The one which is in quadrature with the signal gives phase modulation.

The deviation increases with F_m as has been observed experimentally. This formula has been demonstrated but the result can be guessed if its meaning in terms of carrier/noise-power ratio is understood.

Carrier/noise power ratio is proportional to the quantity $\left(\frac{F_m}{\Delta f}\right)^2$.

The formula shows only that the carrier/noise-power ratio at the input $\left(\frac{C}{N}\right)_{in}$ is degraded by

the noise figure of the amplifier N

the frequency multiplication factor n which multiplies the modulation index,

and is affected by the band-pass frequency characteristic of the system.

5.2. *Practical Examples*

The following examples will show the order of magnitude of f.m. noise due to this process.

Assuming the following figures:

multiplication factor $n = 96$

noise figure $N = 3$ dB

$$\left(\frac{C}{N}\right)_{in} = 120 \text{ dB at } F_m = 100 \text{ kc/s}$$

$$\frac{A(\omega)}{A(\omega_0)} = 1$$

the band pass of the system being wide with respect to 100 kc/s.

Then

$$\Delta f_{\text{r.m.s.}} \doteqdot \frac{nF_m\sqrt{N}}{\sqrt{2\left(\frac{C}{N}\right)_{in}}} = \frac{96\times 10^5\sqrt{2}}{\sqrt{2}\times 10^6} \simeq 10 \text{ c/s}$$

The second example shows an important limitation.

Assume that the power level at the input of the amplifier is 1 mW. The carrier/noise ratio is limited by thermal noise.

The thermal noise power available at room temperature is:

$$P_n = 4\times 10^{-21}B$$

If

$$B = 1000 \text{ c/s}$$

$$P_n = 4\times 10^{-21}\times 10^3 = 4\times 10^{-18} \text{ watts}$$

Then

$$\left(\frac{C}{N}\right)_{in} = \frac{10^{-3}}{4\times 10^{-18}} = 2{\cdot}5\times 10^{14}$$

$$\left(\frac{C}{N}\right)_{in} = 144 \text{ dB}$$

This input carrier/noise ratio leads to a noise deviation $\Delta f = 0{\cdot}6$ c/s at 100 kc/s from carrier.

This result is quite significant taking into account the fact that for various Doppler systems the requirements on the noise frequency deviation are in the 1 c/s range at 100 kc/s from carrier.

5.3. *Power Level Requirements*

The practical conclusion of this part of the study is that the power level at the input of the amplifier must be as high as possible. There appear limitations because the output power of the oscillator is a function of two parameters:

the power gain of the transistor

the maximum power which can be dissipated in the crystal.

Moreover a certain amount of decoupling attenuation is necessary between the oscillator and the first amplifier.

5.4. *Contribution of the F.M. Noise of the Oscillator*

It was shown in Sect. 5.1 that thermal noise in an amplifier produces small amplitude variations and phase jitter. The analysis of the oscillation criterion of a crystal oscillator shows that the phase jitter due to noise which appears in the loop leads to a frequency jitter. The corresponding f.m. noise at the output of the solid-state source has been calculated, using the following parameters.

P_q power dissipated in the crystal

Q_{q0} unloaded Q of the crystal

Q_{qc} loaded Q of the crystal

T absolute temperature

f_0 frequency of oscillation

n frequency multiplication factor.

The simplified h.f. equivalent circuit of the oscillator is given in Fig. 5. The auxiliary networks required for phasing and crystal neutralizing have been omitted.

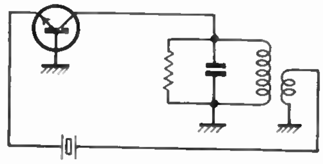


Fig. 5. Simplified h.f. equivalent circuit of oscillator.

Other assumptions are that

the internal feed-back of the transistor is negligible

the output impedance of the transistor is high.

Both conditions are realized for the transistors used in actual operation.

Then at the output of the source:

$$\Delta f = f_0 \sqrt{\frac{KTB}{P_q}} \frac{n}{\sqrt{Q_{q0} Q_{qc}}} \frac{1}{\sqrt{1+4Q_{qc}^2\left(\frac{F_m}{f_0}\right)^2}}$$

The following example will show that the contribution of the oscillator to the total f.m. noise is small. Assuming a relatively poor value of Q_{q0} and Q_{qc}, for instance:

$\sqrt{Q_{q0} Q_{qc}} = 7000$

$f_0 = 100$ Mc/s

$n = 96$

$P_q = 0{\cdot}1$ mW (actually this value could be increased up to 1 mW)

$F_m = 0$ (maximum noise occurring at carrier frequency)

$\Delta f = 0{\cdot}28$ c/s

For $Q_{qc} = 5000$, the spectral distribution is given by the factor:

$$\frac{1}{\sqrt{1+4Q_{qc}^2\left(\frac{F_m}{f_0}\right)^2}}$$

The 3-dB points are 10 kc/s away from the carrier frequency.

6. Experimental Results

The experimental results explained qualitatively in Section 4 will now be compared quantitatively with the theoretical results.

6.1. *Additional Noise*

The linearly-increasing noise deviation gives values around 3·5 c/s in a 1-kc/s bandwidth for an X-band source at 100 kc/s from carrier. Other authorities on low noise systems have claimed results around 1·5 c/s in a 100-c/s bandwidth at the same frequency. These figures are equivalent because the last one has to be multiplied by $\sqrt{10}$ which is the square-root of the bandwidth ratio when compared with results obtained in a 1-kc/s bandwidth. In order to compare these results with the theoretical formula, several noise measurements have been performed on v.h.f. amplifiers. Starting from power levels in the 1 to 10 mW range, low noise v.h.f. transistors (such as 2N918, 2N3287 etc.) have been used in neutralized class A operation. Gain was limited to 13 dB for stability requirements. Disappointing results have been obtained. These transistors give noise figures in the 3 to 3·5 dB range when used in receiver circuits, that is in very small signal operation with optimum generator impedance. Used in medium power amplification, noise figures around 5·5 to 6 dB have been obtained when all parameters were optimized. Optimum conditions are quite critical and practical values as high as 10 dB have been commonly measured.

The noise figure of the second amplifier operating in the 200 mW output power range is not negligible. Typical values are around 20 dB.

If the first amplifier has a noise figure of 10 dB and a gain of 13 dB, the overall noise figure is:

$$N = N_1 + \frac{N_2 - 1}{G_1} = 10 + \frac{100-1}{20} = 15$$

$$N = 12 \text{ dB}$$

Assuming these figures and an input level of 1 mW which is a reasonable value, the contribution of the thermal noise can be calculated and gives:

$\Delta f = 1{\cdot}7$ c/s related to a 1-kc/s bandwidth at 100 kc/s from carrier.

This value is about one half of the overall noise deviation and these figures show that additional noise is a critical parameter in this problem.

6.2. *Oscillator Noise*

The numerical example given in Section 5.4 shows that for values of Q_{q0} and Q_{qc} encountered in practice a crystal oscillator, near carrier, should give a deviation of approximately 0·3 c/s with respect to a 1-kc/s bandwidth. In the same conditions a free-running oscillator having a loaded Q 50 times lower should give a noise deviation of 15 c/s.

In practice, the very low frequency noise components are measured after demodulation in a 10 c/s bandwidth window. Then the figures are multiplied by 10 to be related to the 1000 c/s bandwidth used for higher frequency components. Very near the carrier, actual values are 10 to 20 times greater than calculated values. The actual noise deviation decreases very quickly and approaches the theoretically predicted curve at its 3 dB points.

This observed low frequency excess noise contribution can be explained as follows:

The oscillator noise theory assumes that:

the oscillator circuit in open loop is a linear amplifier

the oscillator is just maintained at the oscillating level

in these conditions there is no possibility of interaction between the v.h.f. carrier and low frequency noise

the cut-off frequency of the oscillating transistor is very high.

Considering the noise figure versus frequency of a typical v.h.f. transistor it is supposed that carrier frequency is lower than the point where noise factor starts increasing with frequency with a slope of 6 dB per octave (Fig. 6). In the frequency domain where the transistor noise figure is flat, the Nielsen formula, which gives its expression, is simplified and:

$$N = 1 + \frac{r_b'}{R_g} + \frac{r_e}{2R_g}$$

r_e, r_b' and R_g corresponding to the equivalent circuit shown in Fig. 7. In these conditions the overall oscillator noise power comes from three sources:

the output impedance of the feed-back network which feeds power to the crystal

the crystal resistance

the input resistance r_{ec} of the transistor (the output being short circuited).

Only these three noise sources have been considered in the oscillator noise theory. If a mixing process exists between carrier and low frequency excess noise, the latter has to be allowed for.

Figure 6 shows that the noise figure of a typical v.h.f. transistor has a low frequency corner around 100 kc/s. The slope of the curve is 3 dB/octave. Starting from values around 4 dB at 100 kc/s the noise figure reaches:

10 dB at 10 kc/s

20 dB at 1 kc/s

and extrapolation to lower frequencies leads to:

30 dB at 100 c/s

40 dB at 10 c/s

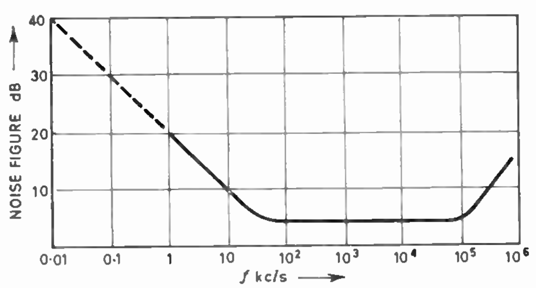


Fig. 6. Noise characteristic of typical v.h.f. transistor.

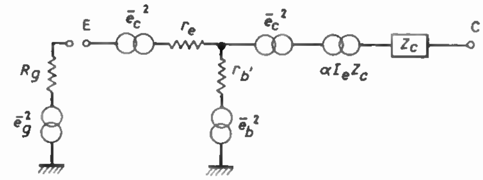


Fig. 7. Equivalent circuit for v.h.f. transistor.

Experimental analysis of the contribution of this mixing process has been started and leads to long and difficult measurements. It needs the comparison of noise spectrum of different types of oscillators, namely, level-controlled oscillators and oscillators operating at lower frequency using transistors having better audio frequency characteristics, etc.

Conclusions on this point have not yet been completed.

However, assuming that mixing process has, for instance, a contribution to the overall noise 20 dB lower than the process due to loop-gain phase-jitter, there is no difficulty in explaining the experimental results.

At 10 c/s the excess noise will be 20 dB above the loop-gain phase-jitter noise.

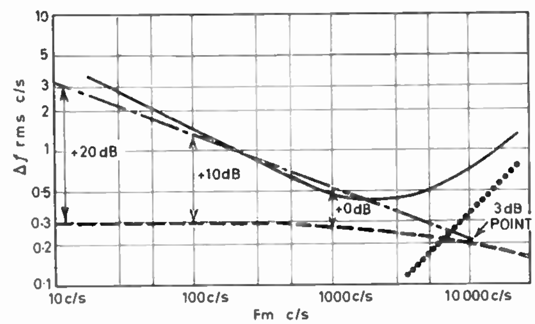
At 100 c/s the contribution will be +10 dB.

At 1 kc/s the contribution will be equal.

At 10 kc/s it is negligible with respect to both oscillator and additional noise. Figure 8 shows that this hypothesis agrees well with experimental results.

6.3. *Contribution of the Varactor Multiplier*

Using the differential phase-jitter measurement equipment, no appreciable noise contribution has been found. The limit of sensitivity of the apparatus being 10^{-5} radian, this means that the multiplier additional noise is lower than 0·1 c/s at 10 kc/s and 1 c/s at 100 kc/s with respect to a 1-kc/s bandwidth.



— — — Theoretical oscillator noise without interaction.

— - — - Theoretical oscillator noise with −20 dB interaction contribution.

· · · · · Additional noise in amplifiers.

——— Noise actually measured.

Fig. 8. Noise performance of oscillator.

7. Methods Developed in order to Improve High Frequency Noise Characteristics

Various methods have been tried in order to reduce the f.m. noise, especially for frequencies away from the carrier. In such applications as slow target coherent Doppler radars, solid-state sources have very satisfactory performances. Their short term stability on time intervals in the 0·1 to 1 ms range leads to radar visibility factors of 50 dB or more which is the limit obtained with good klystrons associated with large cavities and sophisticated stabilizing loops. For fast target Doppler systems and communication systems, the klystron has the advantage that its carrier/noise-power ratio is improved when distance from carrier increases. In order to compete with the best low-noise klystrons in these applications, the carrier/noise-power ratio of solid-state sources which keep a fairly constant value has to be compressed.

7.1. *Compression by Microwave Frequency Discrimination*

In this method (Fig. 9), a small part of the output power drives a microwave frequency discriminator. The output of the discriminator feeds a video amplifier followed by correction networks. The loop is closed by feeding the corrected video signal to a phase modulator.

With such a system which includes phase modulation and frequency discrimination the loop gain increases linearly with frequency. Thus, when compression is applied, the noise frequency deviation becomes flat.

Compression factors of 20 and even 30 dB have been obtained at 100 kc/s from carrier. The use of a microwave discriminator does not increase the size of the source too much. In effect, a relatively low sensitivity discriminator is satisfactory for this use, because, on the other hand, very high sensitivity is obtainable from the phase modulator. Moreover, the modulation index is multiplied in the source by the frequency multiplication factor giving a phase gain. In these conditions, even with a low sensitivity discriminator there is not much problem in obtaining sufficient loop gain using a simple video amplifier which can be realized with two or three integrated circuits.

Microphony is not a critical problem even for tunable systems because the loop gain is very small at low frequencies where mechanical resonances appear.

7.2. *Compression by V.H.F. Phase Discrimination*

Additional noise in the amplifier can be compressed by using a feed-back loop as shown in Fig. 10.

A v.h.f. phase discriminator compares signal phase at the input and the output of the amplifier and the amplified error signal is applied to a phase modulator.

8. Amplitude Noise

The f.m. noise deviation due to additional noise measured in a 1-kc/s bandwidth corresponds to a carrier/noise-power ratio close to 100 dB. Carrier-a.m. noise-power ratio measured in the same condition is in the 115 dB to 130 dB range.

With amplitude modulation the carrier/noise-power ratio is much higher than with f.m. because amplitude noise is not affected by frequency multiplication.

Final results on a.m. modulation in solid-state sources are not yet available.

9. Appendix 1

Relations between Various F.M. Noise Definitions

In this paper, f.m. noise is expressed in terms of carrier frequency deviation Δf in c/s r.m.s. referring to the noise power measured in a 1-kc/s bandwidth rectangular window of centre frequency $f_0 \pm F_m$, f_0 being the carrier frequency.

Then the carrier/noise-power ratio can be written to a first approximation as:

$$\left(\frac{C}{N}\right)_{\mathrm{dB}} = 20 \log \frac{\sqrt{2F_m}}{\Delta f}$$

A usual definition of noise in communications problems considers the noise power in a given bandwidth, for instance a 4-kc/s psophometric window, referring to a signal power level of 1 mW. This noise power P_N is usually measured in picowatts.

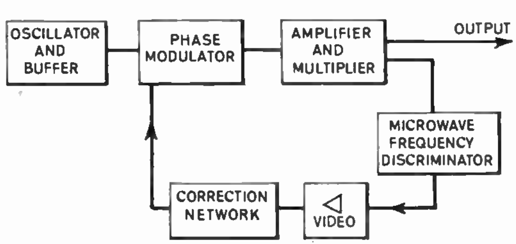


Fig. 9. Compression using microwave frequency discrimination.

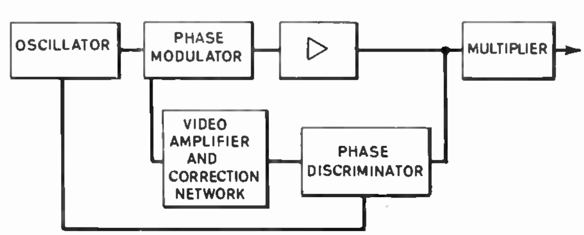


Fig. 10. Compression by v.h.f. phase discrimination.

Then a signal/noise ratio is defined as

$$\left(\frac{C}{N}\right)_{\mathrm{dB}} = 90+10\log\frac{1}{P_N}$$

Considering the carrier/noise-power ratio and referring to a 1 mW level, the relation between Δf measured in a 1-kc/s bandwidth and P_N measured in picowatts in a band B can be calculated:

$$P_N = 5\times 10^5\, C\frac{(\Delta f)^2}{(F_m)^2}B$$

P_N is in pW, C carrier power in mW, B in c/s. For instance if the band of interest is a psophometric window of 4 kc/s bandwidth (equivalent to a 2-kc/s rectangular window).

$$P_N = 10\times 10^8\, C\frac{(\Delta f)^2}{(F_m)^2}$$

Assuming for instance $\Delta f = 1$ c/s r.m.s. at $F_m = 100$ kc/s for a carrier level of 100 mW leads to a noise power in a 4 kc/s psophometric window:

$$P_N = 10\ \mathrm{pW}$$

If the signal/noise ratio has to be related to a given deviation Δf_1 referring to a band B and a 1 mW signal level, then

$$\left(\frac{C}{N}\right)_{\mathrm{dB}} = 20\log\frac{\Delta f_1}{\Delta f}\sqrt{\frac{1000}{B}} - C_{\mathrm{dBm}}$$

C being the carrier power expressed in dB above the 1 mW reference level.

Assuming for instance a reference deviation of 200 kc/s for a 1 mW level, a 4-kc/s psophometric window and a carrier level of 100 mW, if $\Delta f = 1$ c/s

$$\left(\frac{C}{N}\right)_{\mathrm{dB}} = 20\log\frac{2\times 10^5}{\sqrt{2}} - 20 = 83\ \mathrm{dB}$$

10. Appendix 2

Relations between F.M. Noise and Frequency and Phase Short-term Stability

Frequency and phase short-term stability with respect to a given time interval τ can be calculated as soon as the noise spectrum and the demodulation transfer function of the system are known.

The instantaneous angular frequency deviation $\sigma_\omega(t)$ during a time interval τ can be defined as follows.

Assuming that the phase of the signal is at the instant t:

$$\Theta(t) = \omega t+\theta(t)$$

and at the instant $t+\tau$

$$\Theta(t+\tau) = \omega(t+\tau)+\theta(t+\tau)$$

$$\Theta(t+\tau)-\Theta(t) = \omega\tau+\theta(t+\tau)-\theta(t)$$

For a signal of angular frequency ω, the phase accumulation should be $\omega\tau$.

Then the angular frequency deviation during time interval τ can be defined as:

$$\sigma_\omega(t,\tau) = \frac{\theta(t+\tau)-\theta(t)}{\tau}$$

The frequency jitter is due to thermal noise which presents a Gaussian spectral distribution. It can be studied by the calculation of its mean square value

$$\overline{\sigma_\omega^2(t,\tau)} = \frac{1}{\tau^2}\overline{[\theta(t+\tau)-\theta(t)]^2}$$

Following classical development

$$\overline{\sigma_\omega^2(t,\tau)} = \frac{1}{\tau^2}\left[\overline{\theta^2(t+\tau)}+\overline{\theta^2(t)}-\overline{2\theta(t)\theta(t+\tau)}\right]$$

$$\overline{\theta(t)\theta(t+\tau)} = \rho_\theta(\tau)$$

is the correlation function of the phase deviation

$$\overline{\theta^2(\tau)} = \overline{\theta^2(t+\tau)} = \rho_\theta(0) = \int_0^\infty W_\theta(\Omega)\,\mathrm{d}\Omega$$

$W_\theta(\Omega)$ being the spectral density of the phase deviation after demodulation ($\Omega = 2\pi F_m$ being the video angular frequency).

Then

$$\overline{\sigma_\omega^2(t,\tau)} = \frac{2}{\tau^2}[\rho_\theta(0)-\rho_\theta(\tau)$$

It can be demonstrated that

$$\overline{\sigma_\omega^2(t,\tau)} = \frac{2}{\tau^2}\int_0^\infty W_\theta(\Omega)[1-\cos\Omega\tau\ \mathrm{d}\Omega$$

$\overline{\sigma_\omega^2(t,\tau)}$ is null for $\tau = 0$ and $\tau = \infty$, a result which is evident from the definitions used. Obviously, as soon as the frequency noise spectrum is known, there is no difficulty in obtaining the phase spectral distribution after demodulation through a system presenting a transfer function $A(\Omega)$.

If $W_f(\Omega)$ is the frequency noise video spectral distribution

$$W_\theta(\Omega) = \frac{2\pi W_f(\Omega)}{\Omega}A(\Omega)$$

Integration can be performed directly or using graphical methods. Then the quantity $\tau\sqrt{\overline{\sigma_\omega^2(t,\tau)}}$ can be calculated which is the phase jitter on a correlation interval τ and which appears in the computation of sub-clutter visibility factors etc., in various Doppler systems.

Manuscript received by the Institution on 14th June 1965. (Paper No. 1034.)

U.D.C. 621.375.9

Noise Limitation in Helium-Cooled Parametric Amplifiers

By

KURT GARBRECHT,
Dipl.Ing.†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30*th June to* 2*nd July*, 1965.

Summary: The minimum noise temperature of a parametric amplifier depends on the varactor junction capacitance, spreading resistance and its thermal temperature. The time-dependent junction capacitance remains nearly constant by cooling. But the spreading resistance remains constant with cooling only if the varactor is degenerately doped and has an abrupt p-n junction. Actually the junction region is slightly graded and therefore not degenerate. Thus the spreading resistance increases two or three times by cooling from room temperature to 4°K. The higher the thermal resistance of the varactor the more the spreading resistance is heated by the pump power. The measured thermal resistance of some diffused varactors is nearly the same at room and helium temperatures, while the thermal resistance of an epitaxial varactor increases sharply at very low temperatures. So the optimum pump frequency for minimum noise may be much lower than calculated from previous theories which neglect the pump heating. This is because the pump heating decreases with decreasing pump frequency and is of greater importance at low ambient temperatures. Considering the increase and the heating of the spreading resistance the minimum diode noise temperature of a helium cooled parametric amplifier at 4 Gc/s is calculated to be 3°–4°K. This agrees with measured data.

List of Symbols

B bandwidth
C junction capacitance
C_{j0} junction capacitance at zero bias
e unit charge
f_1 signal frequency
f_2 idler frequency
f_c cut-off frequency of the varactor
f_p pump frequency
k Boltzmann constant
n carrier concentration
P pump power
P_{av} available noise power of the spreading resistance
$\tilde{Q}$ dynamic quality factor of the varactor
R_s spreading resistance
$R_{th(e)}$ equivalent thermal resistance
$R_{th(j)}$ thermal resistance with junction heated
S_0, S_1 Fourier coefficient of the time-dependent elastance
T noise temperature of the amplifier
T_a ambient temperature
T_{eq} equivalent noise temperature of the spreading resistance
T_r temperature of the spreading resistance
ΔT equivalent heating
ΔT_0 heating parameter
V_B breakdown voltage
V_0 bias
μ mobility
σ conductivity
ϕ contact potential

† Central Laboratories of Siemens and Halske A.G. Munich.

1. Introduction

The noise of parametric amplifiers is due mainly to the thermal noise of the varactor spreading resistance and can therefore be reduced by cooling the varactor. However the noise temperatures measured for cooled amplifiers exceed markedly the values calculated according to present theories.[1,2,3] An investigation of this phenomenon is rendered difficult by the inaccuracy of measurements of low noise temperatures. The noise temperature T of parametric amplifiers produced by the thermal noise of the spreading resistance depends on the dynamic $\tilde{Q}$ of the varactor, on the idler/signal frequency ratio f_2/f_1, and on the temperature T_r of the spreading resistance. For a high dynamic $\tilde{Q}$ and an idler circuit without additional loading, the noise temperature is determined by the equation[4]

$$T = \left[\frac{1}{\tilde{Q}^2}\frac{f_2}{f_1} + \frac{f_1}{f_2}\right] T_r \qquad \ldots\ldots(1)$$

The dynamic $\tilde{Q}$ of a varactor

$$\tilde{Q} = \frac{S_1}{2\omega_1 R_s} \qquad \ldots\ldots(2)$$

depends on the spreading resistance R_s, the Fourier coefficient S_1 and the signal angular frequency ω_1. S_1 is described by the time-dependent elastance of the diode junction:

$$S(t) = S_0 + S_1 \cos \omega_p t + \ldots \qquad \ldots\ldots(3)$$

ω_p is the pump angular frequency. The temperature T_r of the spreading resistance exceeds the ambient temperature because the bulk regions of the varactor are heated by the pump power.[5]

The noise of a parametric amplifier may further be increased by noise sources in addition to the thermal noise of the spreading resistance. These noise sources are listed in Table 1 in the order of their relative significance.

Table 1

Noise sources in parametric amplifier

1. Thermal noise produced by the spreading resistance heated by the pump power.
2. Thermal noise due to electrical losses within amplifier structure (circulator, input line).
3. Noise due to reflection of noise powers (e.g. of second stage) on the amplifier input.
4. Shot noise of diode currents.
5. Noise of diffusion capacitance.
6. Breakdown noise.
7. Flicker noise
8. Noise due to fluctuations in pump amplitude and frequency.

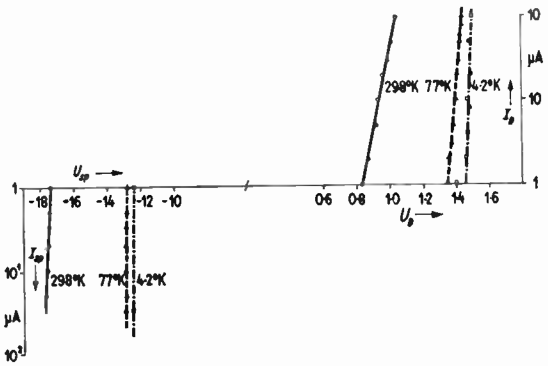


Fig. 2. I/U characteristic of a GaAs varactor at several temperatures.

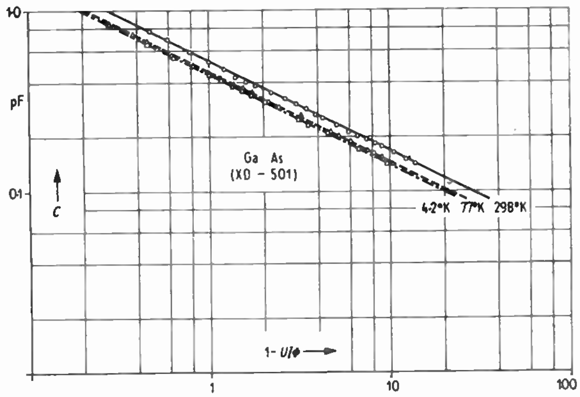


Fig. 1. Junction capacitance of GaAs diode vs. normalized diode voltage.

The effect of the thermal noise caused by the spreading resistance at low temperatures will be discussed. The noise due to electrical losses are well known and the other noise sources listed in Table 1 can be made negligible by proper design.[6]

2. Dynamic $\tilde{Q}$ at Low Temperatures

According to eqn. (2), the dynamic $\tilde{Q}$ depends on the Fourier coefficient S_1 and the spreading resistance R_s.

2.1. *Temperature-dependence of the Fourier Coefficient S_1*

The Fourier coefficient S_1 of the varactor depends not only on the controlled junction capacitance/voltage characteristic but also on the current/voltage characteristic of the diode. In order to prevent shot noise, the amplifier must not be driven in the forward or breakdown region. Figure 1 shows the junction capacitance C of a GaAs diode as a function of the diode voltage V normalized to the contact potential ϕ at various temperatures. The slope of the characteristics, which corresponds to an abrupt p-n junction, is the same at all temperatures. The capacitance decreases slightly if cooling is applied. Figure 2 shows the current voltage characteristic of the same varactor. In the case of cooling, the forward current arises at higher voltages and increases more sharply in agreement with the theory. The breakdown begins at lower voltages. This points to an avalanche breakdown. As can be seen from Fig. 1 and Fig. 2, the time dependence of the junction capacitance and the driving range do not change significantly. As a first approximation, the Fourier coefficient S_1 may thus be considered independent of temperature. The variation of the dynamic $\tilde{Q}$ with cooling is therefore determined primarily by the variation in spreading resistance.

2.2. *Temperature-dependence of the Spreading Resistance*

The spreading resistance of diodes depends largely on the conductivity of the semiconductor in the bulk regions which is given by

$$\sigma = n\mu e \qquad \ldots\ldots(4)$$

The typical temperature-dependence of the conductivity of an ion doped semiconductor is shown in Fig. 3 (solid curve). *At extremely high temperatures* the charge carrier concentration and the conductivity are high due to band-band transitions. *At medium temperatures* the carrier concentration is constant. The conductivity here usually exhibits metallic behaviour because the mobility μ has a negative temperature coefficient.[7] *At low temperatures*, the ionization of impurities decreases, for the activation energy $E_A > kT$). Thus the conductivity decreases rapidly and tends towards zero. This would mean, as has long been assumed, that varactors will no longer function below 20 to 60°K.[8]

However, when the doping of a semiconductor is increased more and more, the discrete energy level of isolated impurities first broadens to an impurity band. With further doping, the impurities merge with the free band. Due to carrier-carrier interactions the carriers are able to move within the impurity band in the same way as in metals. The semiconductor is degenerate. The activation energy becomes zero, the Fermi level falls within an allowed band, and the carrier concentration is temperature-independent. Thus the conductivity is only proportional to the mobility. The mobility in degenerate semiconductors is likewise anomalous and increases only slightly with cooling to about 50°K and remains almost unchanged with further cooling.[7] As shown in Fig. 3 (dashed curve), the conductivity of degenerate semiconductors follows an analogous pattern.

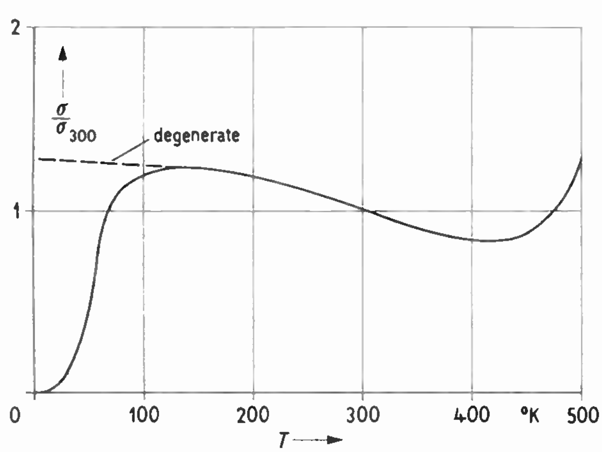


Fig. 3. Electrical conductivity σ of a doped semiconductor vs. temperature. σ_{300} is the conductivity at room temperature.

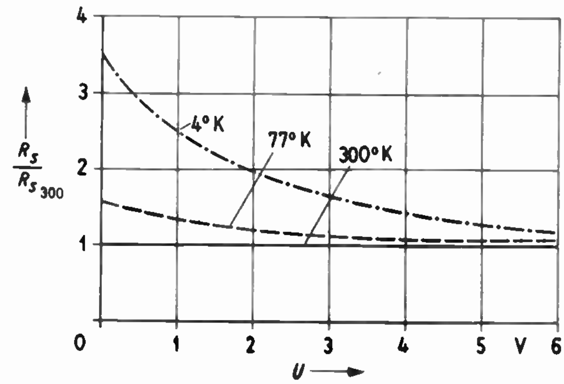


Fig. 4. Spreading resistance R_s of a typical GaAs varactor vs. diode voltage at several temperatures. $R_{s(300)}$ is the spreading resistance at room temperature.

Varactors for parametric amplifiers are mostly degenerate. Thus, in the case of an abrupt p-n junction it is to be expected that, with cooling, the spreading resistance will first decrease and then remain constant. In the case of diodes with a graded p-n junction, on the other hand, the doping is not degenerate in the graded region. With cooling to below 20 to 60°K, therefore the ionization of impurity will decrease here, which means the spreading resistance rises sharply. Moreover, the spreading resistance of cooled diodes with graded junction is voltage-dependent, for the narrower the space charge region, i.e. the lower the applied voltage, the more the low doped graded region with high resistivity will contribute to the spreading resistance. Therefore varactors with a graded p-n junction are unsuitable for cooled amplifiers.[9]

Measurements on varactors with an abrupt p-n junction confirmed these propositions. Figure 4 shows the bulk resistance of a diffused varactor as a function of the applied voltage. Such measurements were first reported by C. Blake *et al.*[10] It is seen from the voltage-dependence of the spreading resistance that the p-n junction must be slightly graded as is frequently the case with diffused varactors. To avoid this, the depth of diffusion must be very limited so as to ensure that the p-n junction will be sufficiently abrupt. On the other hand, the diffusion must be sufficiently deep. Otherwise the p-n junction will be destroyed during the subsequent alloying of the contact. This problem is faced by all manufacturers of varactors. At present users must select suitable diodes by measuring the spreading resistance at 4·2°K.

Silicon diodes are always less suitable for cooled parametric amplifiers because they become degenerate when doped to more than 10^{19} cm^{-3}. With this doping, however, it is only possible to attain breakdown voltages of about 1 V. Silicon varactors can therefore be doped only up to the starting point of degeneration and, with cooling to 4·2°K, a rise in spreading resistance must be accepted. However H. J. Fink *et al.*,[11] report that epitaxial silicon diodes are suitable even at 4·2°K. In the case of GaAs diodes, on the other hand, degeneration starts with a doping of about 10^{17} cm^{-3}. In this case it is possible to realize diodes with breakdown voltages of about 10 V. GaAs varactors can therefore be doped up to the point of complete degeneration.

In short it may be said that the dynamic $\tilde{Q}$ of GaAs diodes remains unchanged with cooling to 4·2°K only if the diodes are doped to the point of degeneration and exhibit a really abrupt p-n junction. Since p-n junctions are more or less slightly graded so far, suitable diodes must be selected by measuring the spreading resistance at 4·2°K. In the case of cooling to 4·2°K it is usually necessary to reckon with about double the spreading resistance.

3. Noise Temperature Considering the Pump Heating

3.1. *Theory of Noise Temperature*

The pump power required for driving a varactor will heat the spreading resistance. The varactor is capable of delivering the available noise power.

$$P_{av} = kT_{eq}B = k(T_a + \Delta T)B \qquad \ldots\ldots(5)$$

T_{eq} denotes the equivalent noise temperature of the spreading resistance and exceeds the ambient temperature T_a by the equivalent heating ΔT. The equivalent noise temperature T_{eq} can be calculated by the integration of the electrical resistance gradient multiplied by the local temperature within the bulk region. In the temperature interval $T_a + \Delta T$ the thermal conductivity in the semiconductor is approximately constant, with the result that the equivalent heating ΔT is proportional to the pump power P,

$$\Delta T = R_{th(e)}P = \Delta T_0 \left(\frac{f_p}{f_1}\right)^2 \qquad \ldots\ldots(6)$$

The proportionality factor $R_{th(e)}$ is the equivalent thermal resistance. The pump power P is proportional to the square of the pump frequency f_p. It is advantageous to define a heating parameter ΔT_0. The equivalent noise temperature of the spreading resistance is increased by ΔT_0 when the varactor is pumped with the signal frequency f_1. The heating parameter ΔT_0 is a characteristic diode parameter like the dynamic $\tilde{Q}$; both depend on the signal frequency f_1 and the pump power.

The noise temperature of the parametric amplifier is, according to eqns. (1) and (6),

$$T = \left[\frac{1}{\tilde{Q}^2}\frac{f_2}{f_1} + \frac{f_1}{f_2}\right]\left[T_a + \Delta T_0\left(\frac{f_2}{f_1} + 1\right)^2\right] \qquad \ldots\ldots(7)$$

Here the pump/signal frequency ratio is expressed in terms of the idler/signal frequency ratio f_2/f_1. The noise temperature becomes a minimum for an optimum idler/signal frequency ratio. This minimum depends not only on the dynamic $\tilde{Q}$ but also on the heating parameter ΔT_0 and the ambient temperature. Figure 5 shows the minimum attainable noise temperature T_{min} and the associated optimum idler/signal frequency ratio $(f_2/f_1)_{opt}$ as a function of the dynamic $\tilde{Q}$ and the relative heating parameter $\Delta T_0/T_a$.[5] The optimum idler/signal frequency ratio decreases considerably with increasing relative heating parameter. This is because the pump heating decreases with decreasing pump frequency and is of more importance at low ambient temperatures.

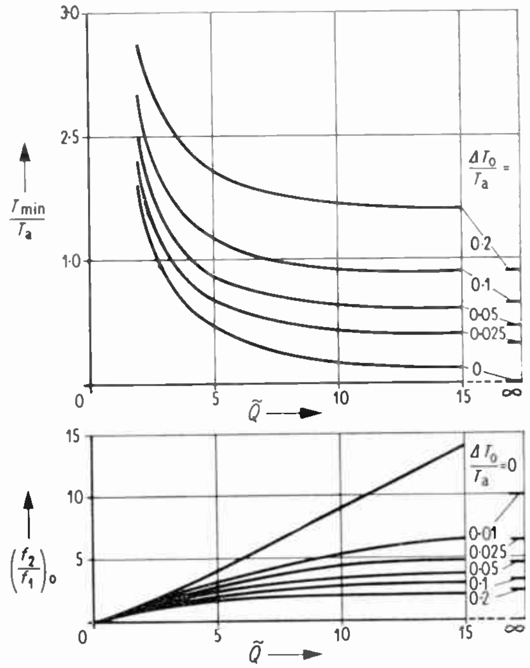


Fig. 5. Minimum noise temperature and optimum idler/signal frequency ratio as a function of dynamic quality factor $\tilde{Q}$ and relative heating parameter $\Delta T_0/T_a$.

As an example, assume a helium-cooled varactor with the dynamic $\tilde{Q} = 10$ and a relative heating parameter $\Delta T_0/T_a = 0{\cdot}1$. The heating parameter corresponds to an equivalent thermal resistance of 420°K/W, if it is assumed that the varactor is pumped at the signal frequency and the pump power is 1 mW. The minimum noise temperature in this case is about 4°K. The optimum idler/signal frequency ratio is only 3 and is nearly independent of the dynamic $\tilde{Q}$. The equivalent noise temperature of the spreading resistance therefore is about $\Delta T = 7$°K higher than the ambient temperature of 4·2°K. If the heating is not taken into account, a theoretical optimum idler/signal frequency ratio of 9 results. In this case the spreading resistance is heated to over 40°K and the noise temperature rises to 9°K.

3.2. *Temperature-dependence of the Thermal Resistance*

The configuration of a diffused mesa diode is shown in Fig. 6. The spreading resistance is located primarily in the n-region (1·5 Ω) with lower doping than the p-doped mesa, and in the contact (0·5 Ω). Therefore heat currents have to flow through the thermal resistance of the n-region (250°K/W), the p-region (1000°K/W) and the wire (2500°K/W).

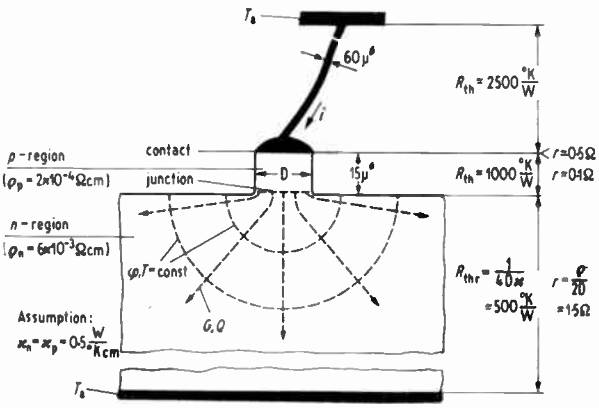


Fig. 6. Cross-section of a diffused varactor with mesa junction.

The thermal resistance is the smaller, the larger the junction area, i.e. the higher the junction capacitance C and the higher the breakdown voltage V_B. The thermal resistance is approximately equal to the product of the reciprocal junction capacitance and the square root of the breakdown voltage,

$$R_{\text{th(j)}} \simeq \frac{1}{C\sqrt{V_B}} \qquad \ldots\ldots(8)$$

Figure 7 showing the thermal resistance $R_{\text{th(j)}}$ between junction and ambient versus junction capacitance and breakdown voltage illustrates eqn. (8). The values for Fig. 7 are obtained from our own measurements and measurements published by Sylvania.[12] The breakdown voltage of varactors used in parametric amplifiers are in the most cases about 6 V. For constant breakdown voltages the equivalent heating according to eqns. (6) and (8) is

$$\Delta T \simeq \frac{f_p^2}{f_c} \qquad \ldots\ldots(9)$$

It can be seen from eqn. 9, that a high varactor cut-off frequency ($f_c = 1/2\pi R_s C$) is also suitable for keeping heating small.

As the thermal conductivity is temperature-dependent,[13] different minimum noise temperatures

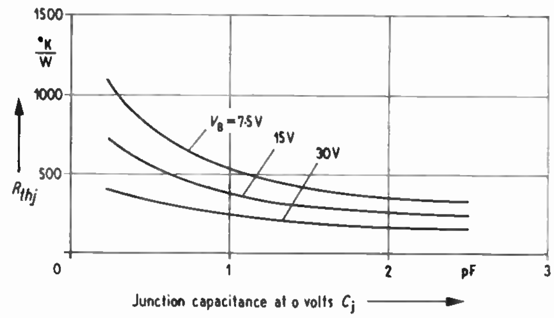


Fig. 7. Thermal resistance of pill varactors measured with a forward current vs. breakdown voltage and junction capacitance.

and optimum idler/signal frequency ratios result for each ambient temperature.

Figure 8 shows the typical temperature-dependence of the thermal conductivity of doped semiconductors. The solid curve represents the average value of the measurements supplied by various authors. At temperatures below 10°K the thermal conductivity is produced by scattering of phonons at the boundary surfaces. The conductivity in this temperature region is therefore dependent on the geometry. Measurements performed so far were on thin cylindrical specimens. Varactors are of the disk-shaped construction and the heat is generated only in a small volume of the large n-region in the vicinity of the junction (Fig. 6). Thus only a minor scattering of phonons will occur at the boundary surfaces and the conductivity below 10°K will consequently be higher than in the measurements stated. This is indicated by a dashed curve in Fig. 8.

These propositions are confirmed by measurements of the thermal resistance of diffused varactors at

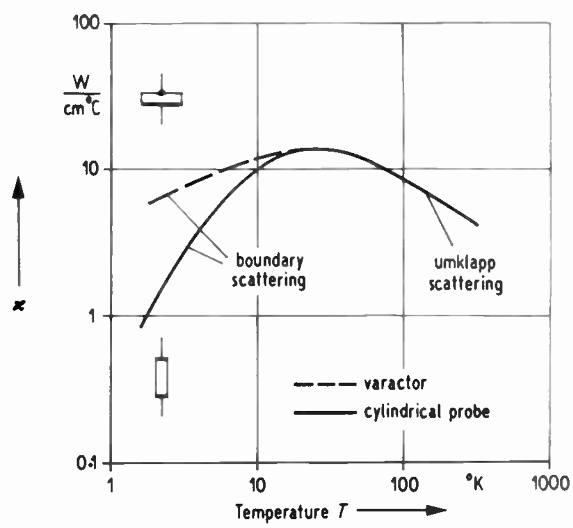


Fig. 8. Thermal conductivity of semiconductors vs. temperature.

several temperatures. The varactor first was heated with a high forward current (e.g. 15 mA or less). After heating the time-dependence of the voltage needed for a small constant current (e.g. 50 µA or less) was measured with an oscillograph (Fig. 9). The ordinate of the oscilloscope photographs can be calibrated directly in degrees Kelvin after measuring the bias for the constant current of 50 µA or less as a function of ambient temperature. The slopes dT/dU of these curves for different diodes are approximately 0·6 deg K/mV at 296°K, 0·8 deg K/mV at 77°K and 1·5 deg K/mV at 4·2°K. The thermal resistance can be calculated from the measured ΔU (Fig. 9) by using

$$R_{\text{th(j)}} = \frac{dT}{dU} \cdot \frac{\Delta U}{P}$$

where P is the pump power.

As can be seen from curves 3 in Fig. 9 a charge-storage effect may occur and must be avoided. The rapid increase of the voltage (curves 2 in Fig. 9) directly after heating is due to the charge-storage effect. At low temperatures the charge-storage effect is more intense. The thermal resistance $R_{\text{th(j)}}$ measured in this way is the thermal resistance between junction and ambient. But the equivalent thermal resistance $R_{\text{th(e)}}$ defined by eqn. (6) corresponds to the thermal resistance between spreading resistance and ambient. As can be calculated from the geometry of the varactor, the equivalent thermal resistance is approximately one-third the thermal resistance measured with a forward current,

$$R_{\text{th(e)}} \simeq \tfrac{1}{3} R_{\text{th(j)}} \qquad \ldots\ldots(8)$$

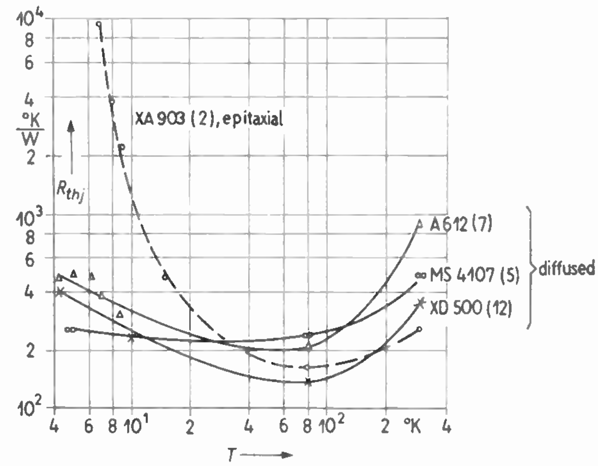


Fig. 10. Thermal resistance of GaAs varactors vs. ambient temperature measured with a forward current.

The thermal resistance measured with some diffused and one epitaxial varactor are shown in Fig. 10. The varactor data are plotted in Table 2. Note the high breakdown voltage of the epitaxial varactor. As to be seen from the high breakdown voltage the epitaxial varactor is not fully degenerate and therefore unsuitable for low cooled parametric amplifiers.

Table 2

Data of the GaAs varactors measured

Varactor	Type	C_{j0} (pF)	V_B (V)	f_c (at − 2 V) (Gc/s)
XA903(2)	epitaxial	0·77	37	250
A612(7)	diffused	0·355	11·8	162
MS4107(5)	diffused	0·455	7·1	255
XD500(12)	diffused	0·48	9·7	98

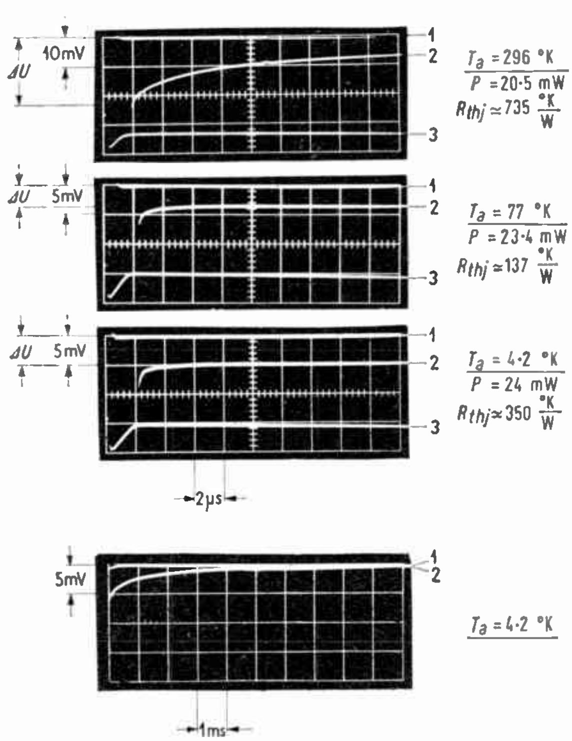


Fig. 9. Oscillograms of varactor (A612) voltage needed for a constant current of 50 µA after heating with a current of 15 mA (curves 2) and without heating (curves 1). Curves 3 plotted with a sensitivity of 0·5 mV per large unit indicates the snap-off effect. The picture at the bottom shows the full time-dependence of the heating.

The thermal resistance of the diffused varactors changes with cooling in a similar manner to the assumed thermal conductivity of varactors (dashed curve in Fig. 7). On the other hand the epitaxial varactor has a thermal response due to the conductivity of this cylindrical specimen (solid curve in Fig. 7). The thermal resistance of the diffused varactors increases only slightly with cooling. Therefore the assumption in the theory of noise, equivalent heating proportional to pump power eqn. (6), is valid. At very low temperature there may be additional heating because of increasing pump power due to an increasing spreading resistance with cooling.

4. Comparison of Measured and Calculated Noise Temperatures

The theory of noise considering the pump heating effect is verified by the noise temperatures of a parametric amplifier measured at various ambient temperatures. The amplifier used is shown on the photograph of Fig. 11. The inaccuracy of the noise measurement is about ± 3 deg K and is further increased by subtracting the noise contributions of the circulator and the input line. The data of the parametric amplifier used are shown in Table 3.

Table 3

Data for the parametric amplifier used with GaAs varactor MS4107

Ambient temperature	T_a	296	77	4·2	°K
Signal frequency	f_1	4	4	4	Gc/s
Pump frequency	f_2	36	36	36	Gc/s
Dynamic Q-factor measured	$\tilde{Q}$	8	8	5·6	
Bias	V_0	−2	−2	−2	V
Spreading resistance	R_s	4·4	4·5	5·8	Ω
Equivalent thermal resistance	$R_{th(e)}$	160	80	90	°K/W
Pump power calculated	P	54	44	55	mW
Relative heating parameter	$\frac{\Delta T_0}{T_a}$	4×10^{-4}	5×10^{-4}	$1{\cdot}5\times10^{-2}$	
Equivalent heating	ΔT	9	4	5	°K
Noise temperature measured	T	83	29	8·5	°K

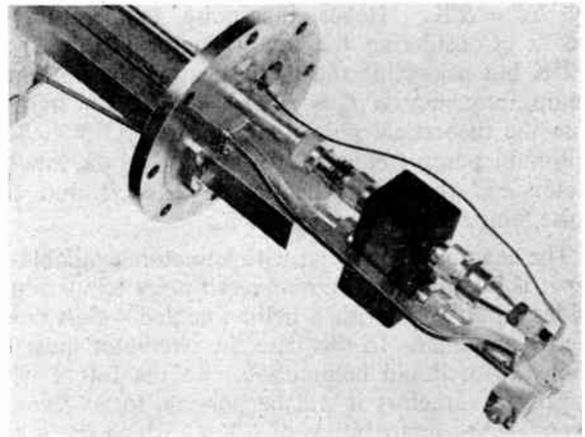


Fig. 11. Parametric amplifier used for noise temperature measurements.

According to Section 3.2 the equivalent thermal resistance $R_{th(e)}$ is assumed to be one-third of the measured thermal resistance $R_{th(j)}$. The pump power needed is calculated from the varactor data, because an unknown part of the pump power delivered to the amplifier was dissipated in the circuit. The pump power seems to be very high, but note the unusually high pump frequency of 36 Gc/s. This pump frequency is optimum if the amplifier is not cooled in liquid helium as is to be seen from Table 4.

As can be seen from Table 4 in the case of helium cooling measured noise temperatures agree with the calculated only if the pump heating is taken into account. The discrepancy between measured and calculated noise temperatures at 4·2°K quoted by other authors can be probably attributed to the pump heating effect.

In the case of helium cooling the pump frequency of the amplifier used is higher than the optimum pump frequency. Thereby the equivalent noise temperature of the spreading resistance is increased

Table 4

Comparison of measured and calculated noise temperatures neglecting heating parameter (A) and taking it into account (B).

Noise temperatures do not include contributions from circulator, input line or second stage.

Ambient temperature	T_a/(deg K)	(A) 296	(A) 77	(A) 4·2	(B) 296	(B) 77	(B) 4·2
Relative heating parameter	$\Delta T_0/T_a$	0	0	0	4×10^{-4}	5×10^{-4}	$1{\cdot}5\times10^{-2}$
Noise temperature measured	T(deg K)	—	—	—	83	29	8·5
Noise temperature calculated	T(deg K)	75	19	2	75	21	5·8
Optimum frequency ratio	$(f_2/f_1)_{opt}$	7	7	4·5	7	7	3·5
Minimum noise temperature	T_{min}(deg K)	75	19	1·5	75	20	2·4

by $\Delta T = 5°K$. Hence the noise temperature is 8·5°K. Considering the low Q-factor (Table 3) at 4·2°K but neglecting the pump heating the optimum pump frequency is $f_p = f_2 + f_1 = 22$ Gc/s. In this case the theoretical noise temperature is 1·5°K. The optimum pump frequency with regard to the low Q-factor *and* the pump heating is 18 Gc/s and the noise temperature becomes 2·4°K.

The example shows that with varactors available so far it is possible to achieve overall noise temperature of less than 10°K with a helium cooled 4 Gc/s parametric amplifier. In this case the circulator must be cooled with liquid helium too. In the future with improved varactors it will be possible to achieve an overall noise temperature of 6°K at which the input line contributes about 3°K, the circulator about 1°K and the varactor 2°K. A second stage parametric amplifier also cooled with liquid helium will not contribute any appreciable noise.

5. References and Bibliography

1. D. C. Hanson, *et al.*, 'Varactor-diode amplifier at liquid helium temperatures', *Digest of technical papers, Intl Solid-State Circuit Conf.*, pp. 54–5, 1963.
2. C. Blake, *et al.*, 'Helium cooled L-band parametric amplifier', *Appl. Phys. Letters*, **2**, pp. 17–19, 1963.
3. M. Uenohara and J. G. Josenhans, 'Liquid helium temperature parametric amplifiers'. Paper presented at the Intl Conf. on Microwaves, Tokyo, September 7th–11th, 1964.
4. K. Kurokawa and M. Uenohara, 'Minimum noise figure of the variable-capacitance amplifier', *Bell Syst. Tech. J.*, **40**, No. 3, pp. 695–722, May 1961.
5. K. Garbrecht, 'Lower limit of paramp noise due to pump heating'. *Digest of technical papers, Intl Solid-State Circuit Conf.*, p. 22, 1965.
6. K. Garbrecht and P. G. Mezger, 'Das Schrotrauschen von Halbeiterdioden in parametrischen Verstärkern', *Frequenz*, **17**, No. 4, pp. 150–7, April 1963.
7. A. K. Jonscher, 'Semiconductors at cryogenic temperatures', *Proc. Inst. Elect. Electronic Engrs*, **52**, No. 10, pp. 1092–1104, October 1964.
8. R. D. Weglein, 'Some limitations on parametric amplifier noise performance', *Trans Inst. Radio Engrs on Microwave Theory and Techniques*, MTT–8, No. 5, pp. 538–544, September, 1960.
9. K. Garbrecht, 'Das Verhalten von Kapazitätsdioden bei tiefen Temperaturen'. Paper presented at the Frühjahrstagung der Deutschen Physik.Gesellschaft in Munich, 1964.
10. C. Blake, *et al.*, 'Varactor-diode behavior at cryogenic temperatures'. Paper presented at the Intl Conf. on Microwaves, Tokyo, September 7th–11th, 1964.
11. H. J. Fink and R. L. Rulison, 'Epitaxial silicon varactors at low temperatures and microwave frequencies', *Proc. I.E.E.E.*, **52**, p. 420, April 1964. (Letters).
12. 'Power Dissipation Capabilities of Sylvania Varactors in Harmonic Generators'. Sylvania Technical Report 9165 G 364.
13. R. O. Carlson, *et al.*, 'Thermal conductivity of GaAs and $GaAs_{1-x}P_x$ laser semiconductors', *J. Appl. Phys.*, **36**, pp. 505–7, 1965.

Manuscript received by the Institution on 3rd May 1965. (*Paper No. 1035.*)

U.D.C. 621.373.51:621.382.232

An Investigation into the Effects of Charge Storage on the Efficiency of a Varactor Diode Doubler

By

B. C. HEAP, B.Sc.(Eng.)†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July, 1965.

Summary: A paper on power dissipation in fourpole networks by A. Weissfloch is adapted to calculate the power dissipation in the input and output matching networks of an experimental varactor doubler and therefore to calculate the true input and output powers from semiconductor bulk material.

Experimentally measured values of input and output power for specified biases are then compared with predicted values for the same biases calculated using an approximate analysis suggested by D. B. Leeson which assumes a non-linear depletion layer capacitance mechanism for harmonic generation.

Reasonable agreement was achieved at power levels and biases which did not involve forward conduction. At higher power levels the results show that if charge storage does occur when forward conduction takes place then it does not enhance the efficiency to any greater extent than could be accounted for by depletion layer non-linearities.

An analysis of the impedance changes of the varactor diode under high power conditions shows that this could account for the decrease in power loss in the surrounding circuits and could well explain the large increases in overall efficiency of the harmonic generator under these conditions.

List of Symbols

C_j	junction capacitance
G_T	maximum available gain of a four-terminal network
I	total a.c. flow through diode
P_{diss}	dissipated power in a four-terminal network
P_{in}	input power to a four-terminal network
P_{out}	output power from a four-terminal network
Q	applied charge to p–n junction
q_{φ}	applied charge at forward voltage φ
Q_b	applied charge at reverse voltage V_b
Q	quality factor
R_s	normalized series resistance of Weissfloch equivalent circuit
R_p	normalized parallel resistance of Weissfloch equivalent circuit
S	p–n junction elastance
V	applied voltage to p–n junction
V_b	reverse breakdown voltage of diode
Y	normalized series reactance of Weissfloch equivalent circuit
Z_g	generator impedance presented to diode
Z_L	load impedance presented to diode
Z_{11} (etc.)	parameters of equivalent four-terminal network representation of a non-linear capacitance
φ	contact potential of a semiconductor p–n junction
Γ	reflection coefficient
γ	exponent of the capacitance—voltage variation

† Formerly in the Department of Electrical Engineering, University College London; now with Ferranti Ltd., Wythenshawe, Manchester, 22.

1. Introduction

Most designers of varactor harmonic generators use the rectifying properties of the diode to self-bias itself to take full advantage of the power handling capabilities of the diode and also design the drive frequency to be usually in the order of 1/10th cut-off frequency of the diode to obtain a reasonably high-Q non-linear reactance.

Ideally, from depletion layer analysis[2] the drive should be allowed to swing into the forward conduction region by an amount equal to the equilibrium potential of the p–n junction depletion layer, φ, to take advantage of the greater degree of non-linearity in charge voltage characteristic. However, this would mean that the diode would rectify, increasing the losses in the diode and hence reducing the efficiency.

If the minority carrier lifetime on either side of the depletion layer is large compared to the period of the drive frequency then as the polarity of the drive voltage reverses, large numbers of carriers tend to be drawn back across the depletion layer without having recombined and hence the net rectified current tends to be reduced. This effect is known as charge storage.

Charge storage has been analysed as a mechanism for harmonic doubling using an idealized charge-voltage characteristic[3] which is a modification of the characteristic by other authors.[4,5] This modification allows the assumption that both input and output circuits can be tuned and the calculated efficiencies compared to those predictions based on depletion layer theory.[6] The calculations show that charge storage may give better performance as a doubler than the depletion layer mechanism.

The effect of the diffusion currents producing charge storage on the efficiency of a varactor multiplier[9] has also been considered and the prediction that the efficiency of a doubler may be enhanced slightly has been made.

If the power dissipation in the coaxial circuits surrounding the diode can be calculated[1,8] then an attempt to compare predictions made, assuming a non-linear depletion layer mechanism for harmonic generation and experimental values can be made. Hence the effect of charge storage on the efficiency of harmonic generation can be evaluated.

2. Power Dissipation in Circuits Surrounding Varactor Diodes

The method of determining the power dissipation in circuits surrounding varactor diodes is based on the technique for determining the Weissfloch equivalent circuit.[1] Weissfloch shows that an arbitrary fourpole network can be represented by an equivalent circuit which is divided up into a lossy part and a lossless part. This renders the equivalent circuit particularly useful for determining the power dissipated in the network. Its adaptation to circuits surrounding varactor diodes is discussed elsewhere.[8]

3. Experimental Analysis of Harmonic Generation

If the power dissipation in circuits surrounding varactor diodes can be calculated then a series of experiments is suggested for determining the effects of charge storage on harmonic generation. As the power input and output from the semiconductor bulk material can be calculated from the harmonic generator input and output powers, a power input/output curve can be plotted under various conditions of bias for the semiconductor bulk material.

The equivalent circuit for the semiconductor bulk material for reverse biases is a linear resistor in series with a non-linear capacitance. The non-linear capacitance variation can be restricted to the characteristic depletion layer capacitance voltage variation given by

$$C_j = \frac{A}{(v+\varphi)^\gamma} \qquad \ldots\ldots(1)$$

and predictions can be made of the output powers using the equivalent circuit assuming the same conditions of input power and bias as were applied in the corresponding experiment. The experimental and predicted curves can then be compared.

The voltage swing can be initially arranged to be completely within the back bias region so that experimental correlation with theoretical prediction should be good because all non-linear effects except that of the depletion layer capacitance should be negligible. However, as the voltage commences to swing into forward conduction, charge storage may begin to occur. The extent to which the experimental curve begins to deviate from the predicted curve at this point should give a good measure of the effect of charge storage on the efficiency of harmonic generation.

The experiments can take two forms. In the first case the bias can be held constant at approximately a half the break-down voltage of the diode and the input power to the semiconductor bulk material gradually raised. The output powers from the semiconductor region can be measured for specific input powers up to the maximum power handling capacity of the diode.

In the second case, the semiconductor region input power can be held constant (not necessarily the harmonic generator input power), and the diode bias varied from approximately a half reverse breakdown voltage towards the region where forward conduction occurs. The output power can be measured for specific biases.

In both these experiments the diodes are initially pumped completely in the back bias region. However, as the input power is increased or the bias decreased the voltage begins to swing into forward conduction and the effects of charge storage can be evaluated as explained previously.

4. Application of Depletion Layer Theory to Harmonic Generation

The following theory for predicting the performance of a varactor diode as a harmonic generator is an adaptation of a method suggested by D. B. Leeson.[2]

4.1. *Taylor Series Approximation*

Figure 1 shows a circuit containing a lossless non-linear capacitor for which the charge-voltage relationship is given by

$$V = V(Q) \qquad \ldots\ldots(2)$$

It is known that

$$V(Q+\delta Q) = V(Q)+V'(Q)\,\delta Q+\ldots \text{etc.} \qquad \ldots\ldots(3)$$

(where $V'(Q)$ is the derivative of $V(Q)$ with respect to charge Q). This is a Taylor series expansion about Q, the derivative being evaluated at the point Q. If δQ is small enough only two terms are required to give a good approximation.

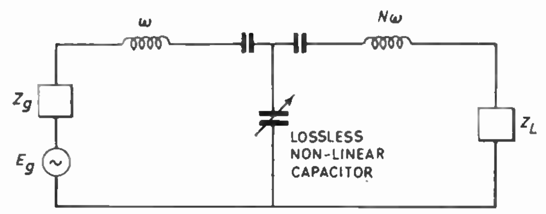


Fig. 1. Circuit for harmonic generator where the currents through the lossless non-linear capacitor are restricted to frequencies ω and $N\omega$.

If $Q+\delta Q$ is a periodic function in time the relationship for the Nth harmonic is given by

$$V_N(Q+\delta Q) = V_N(Q)+\sum_{k=-\infty}^{\infty} V_k'(Q)\,[\delta Q]_{N-k} \qquad \ldots\ldots(4)$$

where δQ is small compared to Q and $V_N(Q+\delta Q)$ is the Nth harmonic of $V(Q+\delta Q)$, etc., i.e.

$$V(Q+\delta Q) = \sum_{-\infty}^{\infty} V_N \exp(\mathrm{j}N\omega t) \qquad \ldots\ldots(5)$$

$$Q+\delta Q = \sum_{-\infty}^{+\infty} Q_N \exp(\mathrm{j}N\omega t)$$

$$V_{-N} = V_N^*, \text{ etc., for real functions}$$

If it is assumed that all currents are zero except the fundamental and Nth harmonic then

$$Q_K = 0 \quad \text{for } K \neq 0, \pm 1, \pm N \qquad \ldots\ldots(6)$$

and choosing

$$\begin{aligned} Q &= Q_0+2Q\cos(\omega t) \\ \delta Q &= Q_N \exp(\mathrm{j}N\omega t)+Q_N^* \exp(-\mathrm{j}N\omega t) \end{aligned} \qquad \ldots\ldots(7)$$

It is seen that Q is an even function in t and therefore $V(Q)$ and $V'(Q)$ are also even functions of t. Hence

$$\begin{aligned} V_N(Q) &= V_{-N}(Q) \\ V_N'(Q) &= V_{-N}'(Q) \end{aligned} \qquad \ldots\ldots(8)$$

and are purely real.

Using equations (4), (7) and (8)

$$\begin{aligned} V_1(Q+\delta Q) &= V_1(Q)+V_{N-1}(Q)Q_N+V_{N+1}(Q)Q_N^* \\ V_N(Q+\delta Q) &= V_N(Q)+V_0'(Q)Q_N+V_{2N}'(Q)Q_N^* \end{aligned} \qquad \ldots\ldots(9)$$

Most functions of $\cos(\omega t)$ decrease monotonically with order and hence the following approximations can be made

$$\begin{aligned} V_{N+1}(Q) &\ll V_{N-1}(Q) \\ V_{2N}'(Q) &\ll V_0'(Q) \end{aligned} \qquad \ldots\ldots(10)$$

Equation (9) can be written as

$$\begin{aligned} V_1 &= V_1(Q)+V_{N-1}'(Q)Q_N \\ V_N &= V_N(Q)+V_0(Q)Q_N \end{aligned} \qquad \ldots\ldots(11)$$

4.2. *Four-terminal Network Representation*

The current through the diode is given by

$$I(t) = \frac{\mathrm{d}Q}{\mathrm{d}t} = \sum_{-\infty}^{+\infty} \mathrm{j}\omega N Q_N \exp(\mathrm{j}N\omega t)$$

and hence

$$I_N = \mathrm{j}N\omega Q_N$$

Equation (11) can be restated in the form associated with four-terminal networks

$$V_1 = Z_{11}I_1+Z_{12}I_N = \left[\frac{V_1(Q)}{\mathrm{j}\omega Q_1}\right] I_1 + \left[\frac{V_{N-1}'(Q)}{\mathrm{j}N\omega}\right] I_N$$

$$V_N = Z_{21}I_1+Z_{22}I_N = \left[\frac{V_N(Q)}{\mathrm{j}\omega Q_1}\right] I_1 + \left[\frac{V_0'(Q)}{\mathrm{j}N\omega}\right] I_N \qquad \ldots\ldots(12)$$

and the behaviour of a pumped non-linear capacitor can be represented by a four-terminal network and its performance can be analysed in a similar fashion.

The maximum available gain of the harmonic generator is given by

$$G_T = \frac{-4R_gR_L(Z_{21}^2)}{|(Z_g+Z_{11})(Z_L+Z_{22})-Z_{21}Z_{12}|^2} \qquad \ldots\ldots(13)$$

Equations (11) and (8) show that Z_{11}, Z_{12}, Z_{21} and Z_{22} are all purely imaginary and hence adjusting the load and generator reactance to cancel Z_{11} and Z_{22} out

$$G_T = \frac{-4R_gR_L(Z_{21})^2}{(R_gR_L-Z_{12}Z_{21})^2} \qquad \ldots\ldots(14)$$

This expression is maximized for

$$R_gR_L = -Z_{12}.Z_{21}$$

and the expression for the conversion efficiency is given by

$$G_T = \frac{-4Z_{21}^2}{-4Z_{21}Z_{21}} = \frac{Z_{21}}{Z_{12}} \qquad \ldots\ldots(15)$$

For a lossless non-linear capacitor, however, the efficiency is unity and therefore

$$Z_{21} = Z_{12} = \frac{V_N(Q)}{\mathrm{j}\omega Q_1} \qquad \ldots\ldots(16)$$

and the first approximation of eqn. (10) is not needed.

4.3. *Adaptation of Analysis to the Lossy Case*

In Fig. 2 the four-terminal network representation for a non-linear capacitor is adapted to account for the additional series resistance of the varactor diode. The resistance, R_s, must be in both the fundamental and harmonic circuits as both fundamental and harmonic currents flow through it in the equivalent circuit of the varactor diode.

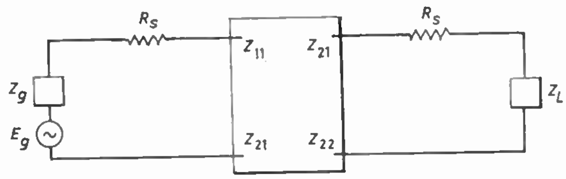


Fig. 2. The equivalent circuit for the varactor diode as a harmonic generator (limited current case).

The expression for the maximum available efficiency of the harmonic generator after adjusting the load and generator reactances to cancel out Z_{11} and Z_{22} is given by

$$G_T = \frac{-4R_g R_L(Z_{21}^2)}{[(R_g+R_s)(R_L+R_s)-Z_{21}^2]^2} \qquad \ldots\ldots(17)$$

This is found to be maximized for

$$R_L = R_g = \sqrt{R_s^2 - Z_{21}^2} \qquad \ldots\ldots(18)$$

and for these values of R_g and R_L the efficiency is

$$G_{T(\max)} = \frac{-4(Z_{21})^2}{(2\sqrt{R_s^2-Z_{21}^2}+2R_s)^2} \qquad \ldots\ldots(19)$$

If

$$m^2 = \frac{R_s^2}{-Z_{21}^2} \qquad \ldots\ldots(20)$$

Then from (19)

$$G_{T(\max)} = \frac{-Z_{21}^2}{-Z_{21}^2(1+2m^2+2m\sqrt{1+m^2})}$$

$$= \frac{1}{(1+2m^2+2m\sqrt{1+m^2})} \qquad \ldots\ldots(21)$$

4.4. *Calculation of Four-terminal Parameters*

To evaluate the performance of the varactor diode as a harmonic generator it is required to Fourier-analyse the function

$$V(Q) = V(Q_0 + 2Q_1 \cos(\omega t))$$

From eqn. (1)

$$\frac{V(Q)}{(V_b+\varphi)} = \left[\frac{V+\varphi}{V_b+\varphi}\right] = \left[\frac{Q+q_\varphi}{Q_b+q_\varphi}\right]^{1/1-\gamma},$$

$$\frac{V'(Q)}{S_{\max}} = \frac{S}{S_{\max}} = \left[\frac{Q+q_\varphi}{Q_b+q_\varphi}\right]^{\gamma/1-\gamma} \qquad \ldots\ldots(22)$$

where V_b is the breakdown voltage of the diode,

Q_b is the applied charge at the breakdown voltage,

q_φ is the applied charge at forward voltage φ.

Therefore

$$\frac{V_N(Q)}{V_b+\varphi} = \frac{1}{2\pi}\int_0^{2\pi}\left[\frac{Q_0+2Q_1\cos\omega t+q_\varphi}{Q_b+q_\varphi}\right]^{1/1-\gamma} \times \cos N\omega t \, d(\omega t)$$

$$= \left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right]^{1/1-\gamma} \times \frac{1}{2\pi}\int_0^{2\pi}\left[1+\frac{2Q_1}{Q_0+q_\varphi}\cos(\omega t)\right]^{1/1-\gamma}\cos(N\omega t)\, d(\omega t)$$

$$\frac{V_N'(Q)}{S_{\max}} = \left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right]^{\gamma/1-\gamma} \times \frac{1}{2\pi}\int_0^{2\pi}\left[1+\frac{2Q_1}{Q_0+q_\varphi}\cos(\omega t)\right]^{\gamma/1-\gamma}\cos(N\omega t)\, d(\omega t) \qquad \ldots\ldots(23)$$

Analysis of parametric amplifiers[9] has shown that the integral

$$I(\alpha, N, \beta) = \frac{1}{2\pi}\int_0^{2\pi}(1+\beta\cos\theta)^\alpha \cos N\theta \, d\theta$$

can be computed by

$$I(\alpha, N, \beta) = \frac{(-p)^N(N-\alpha)}{(1+p^2)^\alpha \Gamma(-\alpha)\Gamma(N+1)} F(-\alpha, N-\alpha, N+1, p^2) \qquad \ldots\ldots(24)$$

where

$$p = \frac{\beta}{1+\sqrt{1-\beta^2}}$$

and

$$F(a,b,c,x) = 1 + \frac{ab}{c}\frac{x}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{x^2}{2!} + \text{etc.}\ldots$$

Therefore for any Q_0 and Q_1 it is possible to compute Z_{11}, Z_{21} and Z_{22} and therefore to determine input, load and generator impedances together with the efficiency and input power.

The input impedance to the varactor is given by

$$Z_{\text{in}} = R_s + Z_{11} + \frac{-Z_{21}^2}{R_L+R_s} = Z_g^* \qquad \ldots\ldots(25)$$

as the generator and load are tuned for maximum power transfer. The load resistance is determined by eqn. (18) and the load reactance by

$$jx_L = -Z_{22} \qquad \ldots\ldots(26)$$

The input power to the semiconductor bulk material is given by

$$P_{\text{in}} = 2|I_1|^2 R_{\text{in}} = 2|j\omega Q_1|^2 R_{\text{in}} \quad \text{......(27)}$$

and the output power by the product of efficiency and input power.

5. Assumptions and Conditions Applied in Obtaining the Experimental and Predicted Curves

Using the London University *Atlas* computer, the values of input, output power, efficiency, bias and Q of the input and load impedances were calculated for the following conditions. $\left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right]$ was varied from 0·05 by increments of 0·05 to 0·5 and β from 0·05 by increments of 0·05 to 0·95 and the above properties calculated for several diodes. The parameters of the diodes defined by eqn. (1) were determined by a method of least-squares curve fitting to measured small-signal capacitance voltage data and the series resistance by methods suggested by Houlding[10] and Mavaddat.[11]

From the above calculated results the series for which $\left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right]$ is held constant at

$$\left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right] = 0{\cdot}5 \quad \text{......(28)}$$

and β is varied from 0·05 to 0·95, is selected and a power input against output curve is plotted (Figs. 5, 7). The experimental curves were obtained by varying the harmonic generator input power until the calculated semiconductor bulk material input power was equal to the specified input power. The bias applied to the diode was that corresponding to the specified input power. The semiconductor bulk material output power was calculated from the harmonic generator output power and with the input power plotted to give points on the experimental curve (Figs. 5, 7).

In this case the power input is varied from low values corresponding to low values of β where the diode is pumped completely in the back bias region to high values where the voltage peak is well into the forward conduction region. This provides the conditions for analysing the contribution of charge storage in varactor harmonic generation as stated previously.

A second set of experiments which also provides these conditions is based on holding the semiconductor region input power constant and varying the bias from high back biases where the diode is pumped completely in the back bias region to low biases where charge storage may again be taking place. The output power is plotted against bias and compared to a theoretical curve calculated assuming the same constant input power.

The calculated curve is determined from the computed results in the following manner. For each separate value of $\left[\frac{Q_0+q_\varphi}{Q_b+q_\varphi}\right]$, the values of β were determined which gave input powers closest to the specified constant input power. As the increments of β are small the variation of input power, output power bias, etc., with β over the increment can be considered linear and therefore the values of output power, etc., corresponding to the specified input power can be interpolated from those corresponding to the determined values of β.

The experimental curve can be obtained as the values of bias applied to the diode are known. For each specific bias the harmonic generator input power can be varied until the semiconductor bulk material input power is that which is specified. The output power from the semiconductor region can be determined from the harmonic generator output power and hence plotted against bias and compared to the predicted curves (Figs. 6 and 8).

6. Calculation of the Forward Conduction Point on the Theoretical Curve

As explained previously, the low powers on the power input, output curves and high biases on the power output, bias curves correspond to diodes pumped completely in the back bias region. It would therefore be an advantage to determine the point on the calculated curves at which the pumping moves into forward conduction. From eqn. (22)

$$\left[\frac{Q+q_\varphi}{Q_b+q_\varphi}\right] = \left[\frac{V+\varphi}{V_b+\varphi}\right]^{(1-\gamma)} \quad \text{......(29)}$$

and therefore

$$\left[\frac{q_\varphi}{Q_b+q_\varphi}\right] = \left[\frac{\varphi}{V_b+\varphi}\right]^{(1-\gamma)} \quad \text{......(30)}$$

Hence the magnitude of charge entailing forward conduction is expressed as a ratio of the total charge stored in the depletion layer at breakdown. The value of β, for any charge biasing $\frac{Q_0+q_\varphi}{Q_b+q_\varphi}$, can now be determined for which the diode begins to be pumped into the forward conduction region.

$$\beta = 1 - \frac{q_\varphi}{Q_0+q_\varphi} = \frac{\dfrac{Q_0+q_\varphi}{Q_b+q_\varphi} - \dfrac{q_\varphi}{Q_b+q_\varphi}}{\dfrac{Q_0+q_\varphi}{Q_b+q_\varphi}} \quad \text{......(31)}$$

If the value of β is known then the input, output powers, etc., can be determined and the point on the predicted curve evaluated.

7. Experimental Harmonic Generator

The previous analysis assumed that the current through the diode was restricted to the fundamental and the required harmonic. The circuits around the diode must therefore ensure that this is the case for any attempt at correlation between theoretical calculation and experimental results to be made.

The fundamental pumping frequency was chosen to be 478 Mc/s. This has a wavelength of 62·8 cm, which is large compared to the physical dimensions of the diode, short circuit and open circuit packages. Therefore any slight mechanical differences in these three packages will have a negligible effect on the position of the plotted impedances. For the same reason the case of the doubler was investigated, the output signal wavelength (31·4 cm) still being large compared to the physical dimensions of the packages. Coaxial circuitry was used to limit the current as the design of such circuitry at this frequency is much easier than that of lumped circuitry.

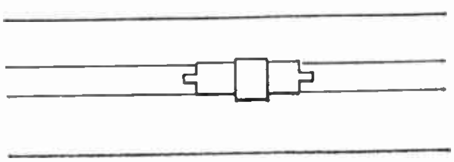


Fig. 3. Diode mount.

The diode was mounted in the coaxial line in series with the centre conductor as in Fig. 3.

Figure 4 shows the coaxial network, suggested by Penfield, set up to conform to the assumptions of the previous theory. Both input and output matching networks consist of two variable short circuit stubs. These matching networks are set to transform the generator and load impedances to those fulfilling maximum power transfer conditions and therefore the assumptions of eqn. (19).

The input filter network consists of a short circuit stub a quarter wavelength long at the fundamental frequency. This presents an open circuit across the main coaxial line for the fundamental signal and a short circuit at the second harmonic. Therefore the fundamental signal is allowed through but the second harmonic is reflected. Similarly the output filter consists of an open-circuit stub one quarter wavelength long at the fundamental frequency. This presents a short circuit across the main coaxial line at the fundamental frequency and an open circuit at the second harmonic. These two filters therefore separate the generator circuit from the load circuit.

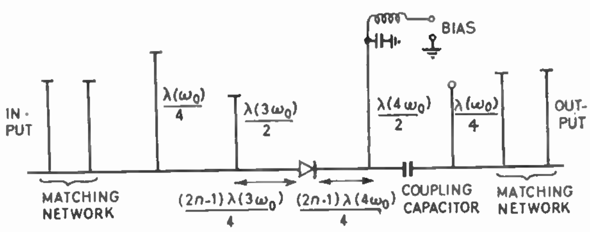


Fig. 4. Experimental harmonic generator.

The circuits for limiting the current through the diode to fundamental and second harmonic consist of a short circuit stub and a low-pass filter stub (to allow bias to be applied to the diode) on opposite sides of the diode mount. The short-circuit stub is a half wavelength long at the third harmonic so presenting a short circuit across the main coaxial line. The low-pass filter stub was of such a length that it presented a short circuit across the main coaxial line at the fourth harmonic. The two stubs are positioned an odd number of quarter wavelengths at the third and fourth harmonics respectively from the plane of the diode.

The diode therefore 'sees' an open circuit at the third harmonic on one side of the diode and one at the fourth harmonic at the other. The currents at these frequencies are suppressed in the diode and assuming that the currents decrease monotonically with order the fifth and sixth harmonics, etc., can be considered negligible. The currents are therefore restricted to fundamental and second harmonic.

The output power from the generator was analysed for the presence of fundamental or any other harmonics to test the effectiveness of the filters used. If any fundamental or other harmonic power was present it was at least 30 dB below that of the second harmonic.

8. Conclusions

Figures 5–8 show reasonable correlation between experimental and predicted curves when the diodes are pumped completely in the reverse bias region (i.e. at power levels below the forward conduction point) and the same degree of agreement is maintained well into the forward conduction region. In the case of Fig. 5 the calculated forward conduction point lies well down the curve as the diode has a breakdown voltage comparable with that of the forward voltage φ. The point at which the diode begins to rectify is at a slightly higher power level than the calculated forward conduction point. However, at these frequencies rectification is not a good indication of power level at the diode and the voltage may be sweeping into the forward region at lower levels but is not exhibited as rectified current due to charge storage. From Fig. 5 it can be seen that if charge storage

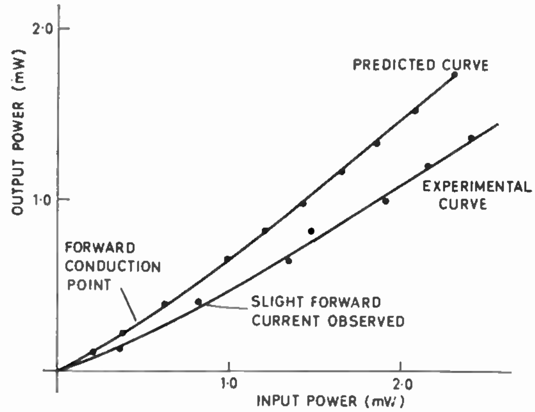


Fig. 5. Power input/output curve for varactor BSC/01 for $\left[\frac{Q_0 + q_\varphi}{Q_b + q_\varphi}\right] = 0{\cdot}5$.

is taking place it does neither enhance nor degrade the efficiency to any noticeable extent. Any increase in efficiency with increasing power level could be accounted for by depletion layer non-linearities only.

Similarly, in Fig. 6 the increase in the efficiency with decreasing bias is only slight and is in agreement with the predictions based on depletion layer theory. The discrepancies in Figs. 5–8 between experimental and predicted curves can only be due to errors in the measurement of the diode parameters or, as is more likely, in the power dissipation calculations as the discrepancies still occur when the diodes are pumped completely in the reverse bias region. The power dissipation measurements for the experimental curves are probably accurate to within ±10%, whilst the predicted curve is probably accurate to within ±5%.

In Fig. 7 there is again reasonable correlation between the experimental and predicted curves at the lower power levels. This experiment was not carried to its conclusion because at power levels higher than indicated on Fig. 7 instabilities began to occur in the system and it was found impossible to maintain the calculated bias. The instability was due to the diode exhibiting a negative resistance in the voltage, current characteristic. It is suggested by Hefni[12] and McDade[13] that this effect is due to a mechanism known as charge multiplication which, however, cannot take place unless charge storage also takes place. This mechanism did not occur in the power output bias experiment of Fig. 8 as the conditions of bias and input power were not suitable for inducing charge multiplication. It seems reasonable to assume that in both the experiments of Figs. 7 and 8 charge storage is taking place when the diode is pumped into the forward conduction region and yet any increase in efficiency can be accounted for by predictions based solely on depletion layer assumptions.

It seems therefore from the above results that although charge storage takes place when the diode

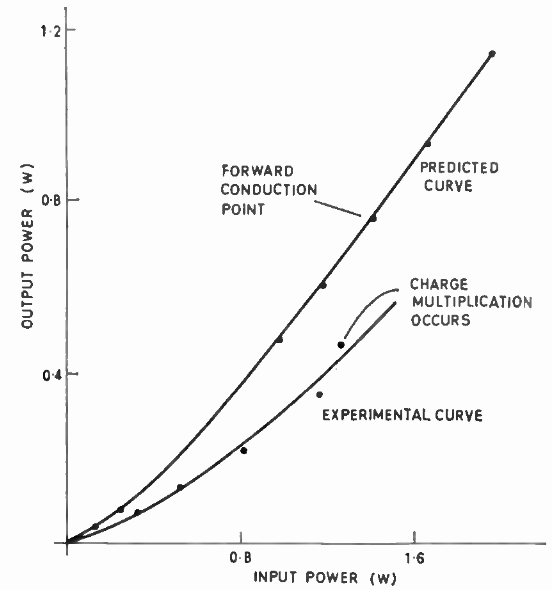


Fig. 7. Power input/output curve for varactor ZC/74/C for $\left[\frac{Q_0 + q_\varphi}{Q_b + q_\varphi}\right] = 0{\cdot}5$.

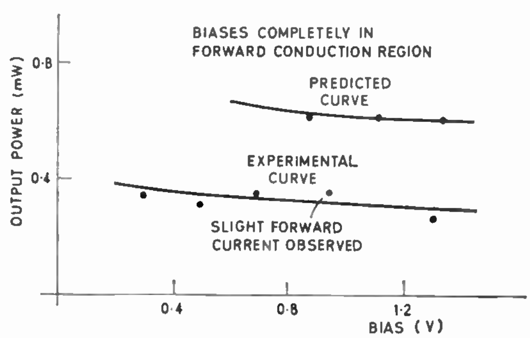


Fig. 6. Power output/bias curve for diode BSC/01 for $P_{in} = 0{\cdot}92$ mW.

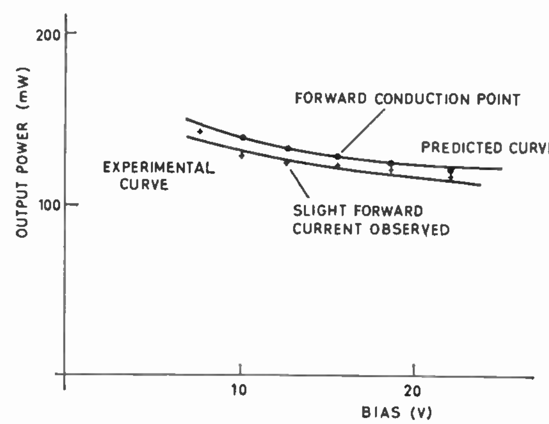


Fig. 8. Power output/bias curve for varactor ZC/74/C for constant $P_{in} = 342$ mW.

is pumped into the forward conduction region it does neither enhance nor degrade the efficiency of harmonic generation to any noticeable extent. The non-linear depletion layer capacitance remains as the basic mechanism of harmonic generation when forward conduction and charge storage take place. However, due to the unavoidably high losses in the experimental harmonic generator, the accuracy of the estimation of the input and output powers is reduced. Increases or decreases in efficiency in the order of a few percent would not be detected. Efforts are therefore being made to design a doubler with low circuit loss to increase the accuracy of the power measurements and to detect small variations in efficiency.

With this experimental doubler it was noticed that the overall efficiency tended to increase with increasing power level or decreasing bias well into the forward conduction region. The variation of power dissipation in the surrounding circuits with power level and bias were studied. In each experiment the fractional power dissipation of the input network generally decreased with increasing power level or decreasing bias. For each specific power level or bias the matching networks had to be reset to obtain optimum conditions for power transfer and therefore satisfy the assumptions of the theory in eqn. (19). A major part of the losses in the input and output networks is probably accounted for by the contact resistance of the variable short circuit stubs and this will vary randomly from one measurement to another. However, a general trend was noticeable and conclusions can be drawn from the results.

It is well known that the fractional power dissipation in any fourpole network is dependent not only on the losses in that network but also on the impedance terminating that network. If a network is terminated by a pure reactance then all the input power to that network is dissipated in that network. However, if a resistance is placed in series with the reactance and gradually increased then the fractional power loss of that network will decrease. A study of the variation of resistance and reactance for the input impedance of the semiconductor bulk material, calculated using eqns. (18) and (25), shows that for the Figs. 5 and 7 the reactance remains constant and the resistance increases for increasing power level. It is also a good approximation for the variation of the input impedance with decreasing bias in Figs. 6 and 8. Therefore from the above analysis the fractional power dissipation in the input network should tend to decrease with increasing power level and decreasing bias.

In Figs. 9, 10 and 11 curves have been drawn of the variation of fractional power dissipation with the calculated input impedance Q and although the random effects mentioned previously are evident the

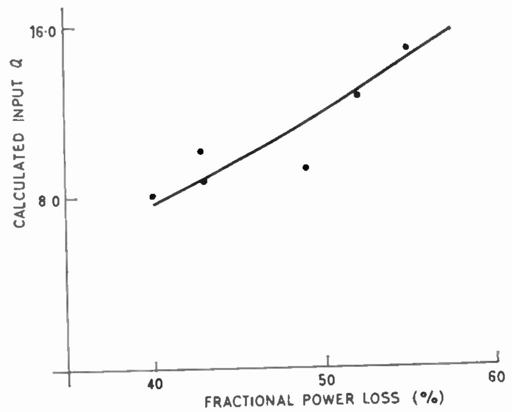


Fig. 9. Power loss against input Q for varactor BSC/01 for $\left[\frac{Q_0 + q_\varphi}{Q_b + q_\varphi}\right] = 0{\cdot}5$.

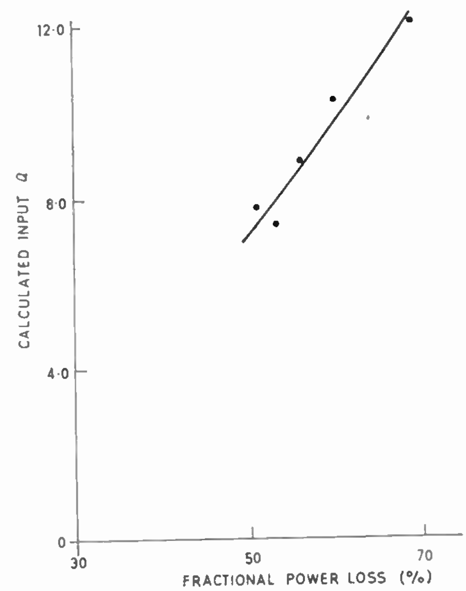


Fig. 10. Power loss against input Q for varactor ZC/74/C for $\left[\frac{Q_0 + q_\varphi}{Q_b + q_\varphi}\right] = 0{\cdot}5$.

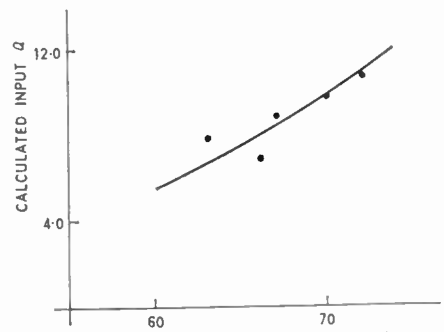


Fig. 11. Power loss against input Q for varactor ZC/74/C for constant $P_{in} = 342$ mW.

general trend of the curves corroborates the above argument. The variation of fractional power dissipation with input impedance Q for the experiment corresponding to Fig. 6 was not drawn as there were insufficient data for the calculated input impedance Q.

Hence for the experimental harmonic generator a culmulative effect takes place. As the input power level is increased or the bias decreased the Q of the input impedance decreases, thereby decreasing the fractional power dissipation in the surrounding circuits. This of itself would increase the overall efficiency of the harmonic generator, however, this decrease in circuit loss increases the input power to the semiconductor bulk material which by the previous predictions based on depletion layer theory increases the efficiency of harmonic generation and the Q of the input impedance. Hence the original decrease in fractional power dissipation is reinforced.

The overall efficiency of the harmonic generator tends therefore to increase with either decreasing bias or increasing power level. Although the losses of the experimental harmonic generator are of necessity high and the above effect consequently magnified, it may occur in harmonic generators whose losses are much smaller.

9. Acknowledgments

The author is grateful to Professor G. D. Sims and Dr. I. M. Stephenson and also the Referees for valuable discussion and suggestions. He is also indebted to Professor H. M. Barlow for laboratory facilities and to G. & E. Bradley Ltd., for the loan of the diodes.

10. References

1. A. Weissfloch, 'Power loss in linear passive fourpole networks', *Elektrische Nachrichtung Technische*, **19**, pp. 259–65, 1942.
2. D. B. Leeson, 'Large Signal Analysis of Parametric Frequency Multiplier', Technical report No. 1701–1, Solid State Electronics Lab., Stanford University, Stanford, California.
3. R. M. Scarlett, 'Harmonic generation with a capacitor exhibiting an abrupt capacitance change', *Proc. Inst. Elect. Electronics Engrs*, **52**, pp. 612–3, May 1964.
4. D. Leenov and A. Uhlir, 'Generation of harmonics and subharmonics at microwave frequencies with p–n junction diodes', *Proc. Inst. Radio Engrs*, **47**, pp. 1724–7, October 1959.
5. D. J. Roulston, 'Efficiency of frequency multiplication using charge storage effect', *Proc. I.R.E.*, **49**, pp. 1961–2, December 1961.
6. T. C. Leonard, 'Prediction of power and efficiency of frequency doublers using varactors exhibiting a general non-linearity', *Proc. Instn Elect. Engrs*, **51**, pp. 1135–9, August 1963.
7. D. B. Leeson, 'Diffusion effects in varactor frequency multipliers', *Proc. I.E.E.E.*, **51**, pp. 1052–4, July 1963.
8. B. C. Heap. 'Determination of power losses in circuits surrounding varactor diodes', *International Journal of Electronics*, **18**, No. 6, pp. 505–27, June 1965.
9. P. Penfield, 'Fourier coefficients of power law devices', *J. Franklin Inst.*, **273**, pp. 107–22, 1962.
10. N. Houlding, 'Measurement of varactor quality', *Microwave J.*, **3**, pp. 40–5, January 1960.
11. R. Mavaddat, 'Varactor diode Q-factor', *J. Electronics Control*, **15**, No. 1, pp. 51–5, July 1963.
12. I. Hefni, 'Effect of minority carriers on the dynamic characteristic of parametric diodes', *Electronic Engng*, **32**, pp. 226–7, April 1960.
13. J. C. McDade, 'R.f. induced negative resistance in junction diodes', *Proc. I.R.E.*, **40**, pp. 457–8, May 1961.

Manuscript first received by the Institution on 15th March 1965 and in final form on 8th May 1965. (Paper No. 1036.)

U.D.C.621.382.3.004.6

Transistors: Reliability, Life and the Relevance of Circuit Design

By

J. M. GROOCOCK, Ph.D., D.I.C., B.Sc., A.R.C.S., C.Eng.†

Originally presented at a Symposium on 'Engineering for Reliability in the design of Semiconductor Equipment' held at Hatfield College of Technology on 13th–14th May 1965 under the aegis of the I.E.R.E. and the I.E.E.

Summary: The paper describes the design features of the highly reliable silicon planar transistors and a brief description of the 'accelerated life test' is given. These tests have to be carried out with high precision. The problem is discussed from the circuit designer's and the manufacturer's point of view and suggestions are made for obtaining reliable performance and a low rate of failure from the transistors.

1. Introduction

From the time of its invention in 1948 the potentially high reliability of the transistor was appreciated. In common with most other electronic components transistors have no moving parts and, therefore, do not suffer from mechanical wear; but, more important, transistors do not operate at high temperatures and, therefore, chemical 'wear out' mechanisms, such as that associated with poisoning of the cathode in a thermionic valve, are relatively unimportant. In principle, therefore, a transistor could go on working indefinitely. However, early transistors were far from reliable. They were not mechanically rugged and the electrical characteristics of the devices could deteriorate markedly as a consequence of changes taking place in the surface of the elements of germanium from which they were made. Reliable transistors only appeared when methods of mounting and encapsulation improved and methods were found to control the ambient atmosphere surrounding the sensitive germanium element. The modern germanium alloy transistor epitomizes this stage of transistor development. The element is inserted inside a hermetically sealed metal can, closed by cold or resistance welding, with the electrical contacts made through glass-metal seals. The gas within the can is kept at constant composition by means of getters such as a molecular sieve.

2. The Silicon Planar Transistor

Modern germanium alloy transistors are extremely reliable but the silicon planar technique allows even higher levels of reliability to be achieved. The silicon semiconductor element enables the device to operate satisfactorily at temperatures more than 100°C, higher than is possible with a germanium device, and to achieve a correspondingly greater power output from a device of given size. Alternatively, for the same performance, the silicon planar device can be drastically 'de-rated' with consequent improvements in reliability. In addition the silicon surface instead of being sensitive to any changes of ambient gas is protected by a stable, impervious layer of silicon oxide, the 'planar' layer. For the silicon planar transistor the conventional hermetically sealed metal can is a second line of protection rather than an essential feature of the device as is the case for a germanium transistor.

For these reasons the silicon planar transistor can have extremely stable characteristics and very high reliability. Figure 1 shows the results of an operational life test at maximum rating (300 mW) on 100 BSY 27 (CV7431) silicon planar epitaxial transistors over an eight-month period, and also the results of a storage life test at 100°C on 500 BSY 27 transistors, for the same period. Similar results are given in detail in Ref. 1. The characteristics show virtually no change during the life test.

The results given in Fig. 1 are from a life test involving 3·6 million transistor hours. The direct cost of such a test in devices, measuring and reporting, is of the order of £500. The capital cost of the conditioning and measuring equipment, which could, of course, be used for other tests, might be between £5,000 and £25,000 or even more depending upon whether a simple manual system or an automatic system was used. Even with no failures all that this test has achieved is an assurance of a failure rate of 0·4% per 1000 hours under operation, 0·08% for storage and 0·06% per 1000 hours if it is assumed that the two results can be added (each figure is given at 90% confidence). Additionally, there is evidence that the characteristics are very stable

† Standard Telephones & Cables Transistors Ltd., Footscray, Kent.

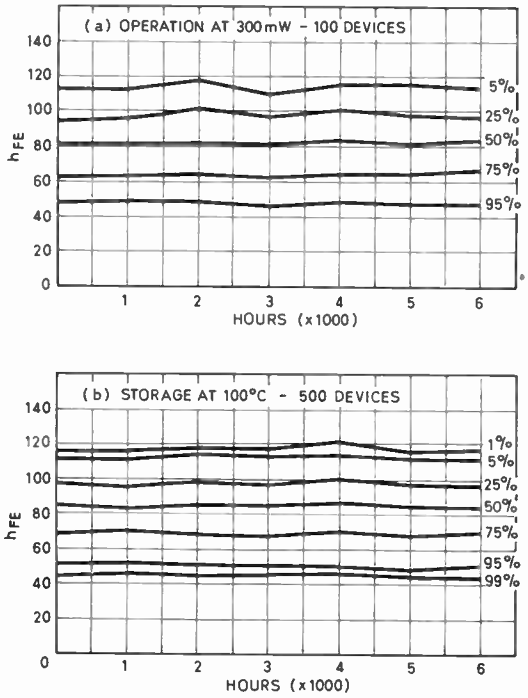


Fig. 1. Percentile plots showing the change of d.c. gain of BSY 27 (CV 7431) transistors life tested for 6000 hours.

so that establishment of tight degradation failure criteria would not cause an increase in the failure rate.

The failure rate assured is an order more thar needed by many users and yet the cost of testing alone is such that it could only be borne economically if orders for very large numbers of transistors were involved. For an order of 100 000 devices the life test described above would add a cost of one or two pence to each device. In addition, for the statistical validity to be properly maintained, it would be necessary to extract the life test sample from the 100 000 devices and to store the remainder for almost a year while the test was carried out.

3. Accelerated Life Testing

Because of the very high costs involved in life tests which simulate actual operating conditions there is great pressure to find methods which reduce these costs. The use of accelerated life tests or over-stress tests is one of the most fruitful approaches. The theoretical basis of these methods and the practical application of them have been described in detail elsewhere.[2] What can be achieved can be illustrated by imagining that instead of life testing the BSY 27 transistor in the conventional manner described above, a similar amount of money had been spent on an accelerated life test, using a similar number of devices. The conventional test indicated with a 90% confidence that the population of devices should have a failure proportion of no more than 0·4% during a 6000 hour period. By dividing the 600 devices into two samples of 300 and carrying out an accelerated test it would be possible to assure a failure proportion of 0·8% not for 6000 hours but for the whole life of an equipment, say 20 years. The assured failure rate is now

$$(0{\cdot}8 \div 20 \times 8{\cdot}57) = 0{\cdot}005\% \text{ per 1000 hours,}$$

rather than the 0·06% given by the conventional test. The confidence level, statistically, is still 90% but practically it must be lower for the accelerated test, because the assumptions implicit in accelerated testing do not have a 100% certainty of being correct.

It is worthwhile emphasizing a number of practical points which are highlighted by this comparison of conventional and accelerated life tests:

(*a*) The cost of the accelerated test was no less than that of the conventional test. It is not true that serious and useful accelerated life tests can be carried out without the expenditure of appreciable amounts of money.

(*b*) The numbers of devices tested in the two cases was the same and yet the conventional tests assured a lower failure proportion (0·4%) than the accelerated test (0·8%). There is another technique, based upon the assumption that the failure *distribution* in time can be extrapolated, which can be grafted on to the accelerated test method and which can be used to give an assurance of low failure proportion. However, the validity of this method is questionable.[2]

(*c*) The accelerated test covers 'early' failures, 'wear out' failures and (because the sample sizes are adequate to assure statistically the required failure proportion) 'rogue' failures. The coverage of rogues is only partial because there cannot be certainty that all of the rogues present in the sample have been revealed by the accelerated test.

(*d*) The main advantage of the accelerated method was in time compression. A prediction of 20-year behaviour was made on the basis of the test. A consequence of this was an apparent marked reduction in the assured failure rate from 0·06% per 1000 hours to 0·005% per 1000 hours. The reduction in this case was by a factor of 12 which is, of course, a very great improvement. It is not, however, reasonable to expect accelerated testing to give failure rate reduction factors of 1000 or more with any useful validity.

To emphasize the point about the cost of accelerated testing, an advanced system which has been built for carrying out life testing on silicon planar transistors for use in submarine repeaters will be described briefly.

The required low failure rate during a 20-year period can only be assured by accelerated test methods. The equipment enables any desired electrical operating condition to be applied to 1600 transistors and simultaneously for them to have their cases clamped at any temperature up to 400°C. For measurement the oven doors to which the heat sinks containing the transistors are bolted are transported to measuring chambers which are accurately temperature controlled. Seven characteristics are measured on each transistor, the results being recorded directly on to punched cards. Scanning of the seven characteristics and through groups of 47 transistors is fully automatic. A full account of the operation of this equipment is to be published shortly.[3]

4. Investigation of the Deterioration of the Internal Connections of a Silicon Planar Transistor by Over-stress Methods

In an earlier section the design features of the silicon planar transistor which make it especially reliable were described. However, achievement of the potentially high reliability is by no means straightforward and great experience with the technique of making planar transistors is necessary. The silicon planar transistor has one reliability problem in common with other high frequency transistors: that associated with the necessity of making contact to very small areas. The fine wire interconnections which are required as a consequence can be a source of unreliability. The difficulty is increased because possibly the easiest method of making this connection is to evaporate aluminium on to the silicon to make the initial electrical contact and then to attach a fine gold wire to the aluminium by thermocompression bonding.

Unfortunately, gold and aluminium form a number of alloys of very low mechanical strength and in the presence of silicon the rate of formation of these alloys (the so-called purple plague) is rapid at temperatures between 200 and 300°C and quite significant, where long life is needed, at temperatures well below 200°C. The use of techniques such as ball bonding which give a high initial strength is a palliative but undoubtedly the best system from a reliability viewpoint is to use aluminium instead of gold wire, eliminating the possibility of plague formation.

An account of an extremely extensive evaluation of the mechanical ruggedness of an aluminium wire bonded silicon planar transistor, the BSY 27, has been recently published.[4] This evaluation involved many thousands of transistors and a very wide variety of mechanical and thermal over stress tests. In this section the results of some more recent experiments will be given.

In order to observe the rate of deterioration of the strength of the internal bonds of silicon planar transistors under high temperature thermal treatment the 'step stress' technique was used.[2] Two samples, each of 400 transistors selected from the same population, were subjected to a 'tumble' test in which they were dropped 100 times on to a steel plate. At each drop the devices suffered shocks of approximately 20 000 g. No device went open circuit under this test.[4] The samples were then heated at 200°C for 20 hours. To determine if this thermal treatment had caused any serious weakening of the bonds the devices were again subjected to the 100-drop 'tumble' test and were then again tested for open circuits. This procedure was repeated, using progressively higher temperatures in the series 200°C, 240°C, 280°C,. . . , 440°C. The whole experiment was then repeated on a further 800 devices taken from the same population but using a 160-hour storage period. The results obtained are given in Table 1. Progressively larger numbers of failures are produced as the temperature is raised from 240°C to 400°C. The first point to note in detail is the very good agreement obtained between the results of the identical experiments on Lots A and B; the small differences are within the range to be expected from sampling errors. Similar conclusions apply to the

Table 1

Results of step-stress experiments on silicon planar transistors

20-hour period at each temperature, 100 drops

Temperature °C	Cumulative failures from Lot A = 400	Cumulative failures from Lot B = 400	Total failures A+B
200	0	0	0
240	0	1	1
280	2	3	5
320	23	23	46
360	70	69	139
400	186	194	381
440	263	265	528

160-hour period at each temperature, 100 drops

Temperature °C	Cumulative failures from Lot C = 400	Cumulative failures from Lot D = 400	Total failures C+D
200	1	0	1
240	2	0	2
280	15	15	30
320	30	25	55
360	35	34	69
400	150	161	288
440	271	263	534

results for Lots C and D. The results given in Table 1 were obtained, in fact, from the third attempt to carry out this experiment. In the first two experiments small errors in experimental technique lead to results with poor agreement for ostensibly identical conditions. The differences were greater than could have been given by sampling errors and were also as great as those produced by the deliberate change from 20 hours to 160 hours. The results were such that it was impossible to make extrapolations.

Having got good reproducibility the cumulative number of failures was plotted on probability paper against the reciprocal of the absolute temperature (Fig. 2). It is apparent that the failures are approximately normally distributed with regard to the reciprocal of absolute temperature. However, the effect of the change from 20-hour steps to 160-hour steps is far from clear. One interpretation of the results would be that the devices are weakened by quite a brief period at any particular temperature and that changes in time within the range 20 to 160 hours have no significant effect. If this interpretation were correct it would imply that less than 0·75% of the devices (at 90% confidence) would be seriously weakened by heating for any periods at temperatures up to 200°C.

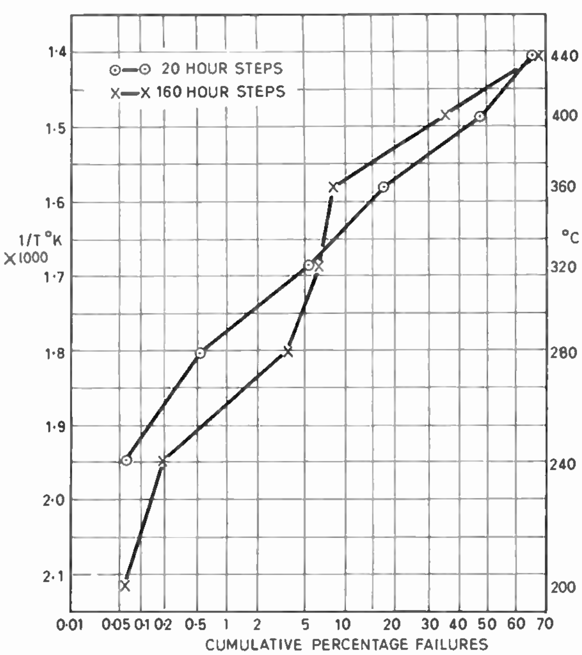


Fig. 2. Plots of cumulative percentage failures against temperature expressed as $1/T$°K for two samples of 800 BSY 27 transistors subjected to step-stress temperature treatment plus tumble testing (100 drops) after each step.

There is, however, some indication, particularly for failure proportions below 5%, that the 160-hour steps are more effective in producing weakening than the 20-hour steps. This is, in fact, the result that might have been expected if the Arrhenius equation, which has wide applicability in reactions of this kind,[2] had been followed. By applying the Arrhenius equation to the failures below 5% it is possible to calculate the expected proportion of devices to be seriously weakened by any period of storage at any temperature.

One result obtained from such a calculation is that 0·2% of the devices would be weakened by 100 000 hours (11½ years) at 170°C. Despite the very good agreement obtained by repetition of identical tests it is the writer's belief that the fact that no firm quantitative conclusions could be made from the results of this experiment was not because of any intrinsically complex failure pattern by the transistors but was a consequence of experimental irregularities, probably for the 160-hour tests at 360°C and above. It is very likely that repetition of the experiment with even greater care would resolve these difficulties. However, it is doubtful if in practice this would be justified in the face of the strong qualitative indication that the particular failure mechanism revealed by this experiment is unlikely to be significant except at high temperatures.

The experiment described above emphasizes an important point about accelerated life tests which the writer believes has general validity. This concerns the high precision with which such tests, if they are to be successful, must be carried out. In this particular experiment it was only at the third attempt that statistical reproducibility was obtained and even then it was not possible to draw firm quantitative conclusions, probably because of loss of experimental control.

Examination of Fig. 2 clearly indicates that relatively small changes in the experimental points would substantially alter the overall pattern of results and markedly affect the extrapolated predictions. Conventional life tests, particularly those of the kind that appear in military specifications, are insensitive to changes in operating conditions. Applying military specification writing to this particular experiment one might call for an AQL of 0·4% for a 20-hour life test at 200°C. Both samples A and B easily pass this test. Even if by mistake the oven temperature was allowed to get up to 240°C and the devices were left in storage for 160 hours the samples would still pass the test (the acceptance number for a 0·4% AQL and 400 devices is four rejects). Changes very much smaller than this would completely invalidate almost any accelerated life test, in which a serious attempt was being made to predict by extrapolation the reliability under realistic conditions.

5. Reliability Clauses in Military Specifications

For many years reliability clauses have been included in military specifications for semiconductor devices. The British Joint Service Specifications in the series CV 7000 are typical of these. The tests appear mainly in Group B, but tests are also sometimes put into Group C. Table 2 shows the Group B specification of the CV 7430–31.

Another part of Group B, which is not illustrated, defines the conditions for leakage current and gain which have to be satisfied after the performance of each of the Group B tests. The main reliability clauses are given in Sub-Groups 7 and 8, although other clauses covering such tests as temperature cycling and vibration fatigue are obviously relevant to reliability. For a typical lot size of 5000 the manufacturer is required to perform a 1000-hour storage life test on 150 devices and a 1000-hour operating life test on 35 devices. The manufacturer with a good product can earn both reduced inspection and reduced duration in which case the amount of testing required is less than one tenth that for normal inspection and duration. It was pointed out in the first section that even a life test extending to several million transistor hours could not assure reliability at the level often required. The military specification life tests which are an order of magnitude smaller can certainly not, therefore, assure good levels of reliability in a statistical sense. The fact that the tests are carried out at maximum rating conditions and are, therefore, accelerated tests with regard to most normal operations of the devices, may be some help. However, the tests are in no sense properly conducted accelerated life test experiments from which quantitative extrapolations can be made.

Table 2

Group B Inspection. Tests of Joint Service Specification CV7430–31

Examination or test	Test conditions		AQL %	Insp. level	Symbol	Limits		Units
	K1007/NATO Ref.	Specific conditions				Min.	Max.	
Sub-Group 1			6·5	IC				
Physical dimensions	5.1	According to drawings 10.3.2.4 and 10.4.2.4						
Sub-Group 2			4·0	IA				
Solderability	5.13							
Temperature cycling	5.5	−55°C to +175°C						
Moisture resistance	5.3							
Sub-Group 3			4·0	I				
Vibration fatigue	5.15.1							
Sub-Group 4			6·5	IA				
Lead fragility	5.10.2	2 cycles						
Sub-Group 5 Omitted								
Sub-Group 6 Omitted								
Sub-Group 7			4·0	I				
High temperature life (Non-operating)	6.2.1 6.6.1.2.2	T_{stg} + 175°C Duration 1000 hr.						
Sub-Group 8			4·0	IA				
Operating life	6.3	V_{CB} = 15V. Duration 1000 hr.						
	6.6.1.2.2	T_{amb} may be at any single temperature between +25°C and +125°C with P_{tot} corresponding to that given on the derating curve						

The problems of assuring reliability are well appreciated by the writers of military specifications and attempts are being made to make the specifications more efficient. One method, which plays an important part in the proposed new Common Standard Specifications, places emphasis upon the accumulation and publication of the results obtained from tests on a whole series of released lots. In this way, provided methods of production can be held constant, considerably higher levels of reliability can be assured over a period of time than can be obtained from single lot results. Another approach is to introduce explicitly accelerated tests as part of the specification. The difficulties associated with this have been touched upon in earlier sections.

6. Transistor Reliability Information Devised from Equipment Usage

In volume, the largest potential source of transistor reliability information is contained in actual equipment usage. There are great difficulties in exploiting this source but nonetheless in two different respects considerable advances have been made. Firstly, a number of equipment manufacturers have published detailed accounts of the reliability of particular components in specific equipments. Cox and Rankin have given a very good example of this sort of work.[5] Secondly, various bodies produce documents giving the integrated results of reliability information derived from a large number of different equipments. In this country workers at the Royal Radar Establishment have taken a leading part and a number of different groups are doing similar work in the U.S.A. Figure 3 is extracted from such a compendium prepared by an American company.[6] It is based upon 3929 million transistor hours of operation in equipments. This is, of course, a volume of testing that would be impossible to achieve by experimental life test. The figure relates the expected failure rate to the power dissipation (relative to the rated power) and the ambient temperature (relative to the rated maximum operating temperature and 25°C). Failure rates vary from 0·02 to 0·2% per 1000 hours. Failures are defined as open and short circuits and radical departures from initial characteristics.

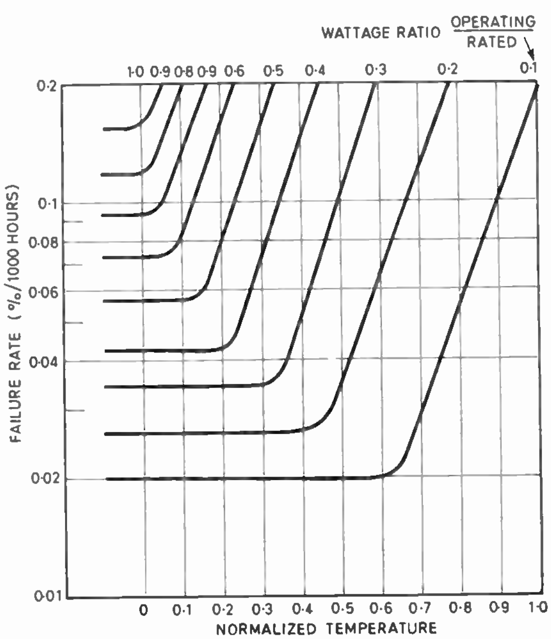


Fig. 3. Transistor failure rates as a function of operating power and ambient temperature. Derived from equipment operation results. (After Ref. 6.)

Summaries of this kind, although backed by very large quantities of information, are of only limited usefulness. They show what may reasonably be expected from most good quality devices—and even for this they usually reflect the characteristics of devices made several years ago rather than current products—but give no direct information at all about the behaviour of particular types of devices in particular equipments.

7. The Position of the Circuit Designer

The previous sections have indicated the complex nature of reliability assurance for transistors. What should the circuit designer do? The answer to this question will depend upon how serious is his interest in reliability. Where reliability is of key importance active co-operation between the equipment maker and the transistor producer is required. Highest reliability will be achieved if the transistors are made and inspected with special care and if the circuit conditions allow substantial de-rating of the devices for temperature, power and voltage. Mechanical stresses due to shock, vibration and temperature cycling should be minimized, and the circuits should be tolerant of changes in characteristics. Coupled with the attempt to make highly reliable devices must be a test programme to give assurance that the required reliability has been achieved. The programme will involve an elaborate series of normal and over-stress tests covering resistance to mechanical, thermal and electrical stresses. By these methods it is practicable to give assurance of failure proportions as low as 0·2% in 20 years, equivalent to failure rates of 0·001% per 1000 hours. However, the cost of the test programme alone will be in excess of £10 000 and the cost of making the devices must be added to this. The capital cost of setting up a facility capable of testing in this manner if this is not available will, of course, be several times £10 000. The duration of the test programme is likely to be in excess of one year.

The equipment manufacturer who needs good reliability but cannot pay for an exercise such as has just been described will markedly improve his chances of achieving this if the following points are borne in mind and he is prepared to pay a reasonable price premium for them.

(*a*) Establish a reliability requirement that is within the 'state of the art'.

(*b*) Select a transistor made by a technology known to be capable of giving high reliability.

(*c*) Buy the transistors from a manufacturer with substantial experience of the technology.

(*d*) Insist on seeing the manufacturer's quality control, reliability activities, and life test results. The latter will certainly not refer to the actual lots to be supplied; they are most unlikely to be on a scale to give the required assurance; they will not reflect the user's actual condition of operation, and they may not even refer to the actual type to be purchased. However, they should indicate the level of characteristic stability likely to be achieved.

(*e*) Buying should be done to a specification which includes quality and reliability clauses backed by lot testing. CV 7000 specifications and, in future, Burghard specifications set the pattern.

The above procedure will not give a statistically assured guarantee of a particular low failure rate. It will, however, ensure that types of devices with a low probability of having a good reliability performance will be eliminated and it will indicate to the transistor manufacturer that the customer is seriously interested in high reliability and is willing to pay a reasonable amount to increase his chances of getting it.

The lowest level of reliability requirement comes from the equipment manufacturer who says he needs very high reliability but is not prepared to pay anything for it. He buys devices at the lowest price he can get, from whoever will sell them. All that he can do is keep his fingers crossed. He may well be lucky. Standard commercial transistors, particularly silicon planars, can have very high reliability and it is by no means unusual for them to give failure rates of 0·01% per 1000 hours in equipment usage.

8. References

1. Life Test Bulletins, Nos. 1 to 5, Standard Telephones & Cables Ltd., Semiconductor Divison.
2. J. M. Groocock, 'Accelerated life testing and over-stress testing of transistors', *Electronics Reliability & Micromin.*, **2**, No. 3, pp. 191–204, December 1963.
3. J. Bickley, 'Over stress life tests on silicon planar transistors using an automatic life testing installation'. To be published.
4. J .M. Groocock, 'Finding the reliable transistor—mechanical and thermal testing of silicon planar transistors', *British Commun. Electronics*, **12**, No. 7, pp. 429–33, July 1965.
5. P. Cox and K. F. Rankin, 'Reliability in large electronic data processing systems', *A.T.E. Journal*, **21**, No. 4, pp. 162–77, October 1965.
6. 'Part Failure Rate Data'. International Telephone & Telegraph Corporation, 1st June 1963.

Manuscript first received by the Institution on 14th May 1965 and in final form on 15th October 1965. (Paper No. 1037.)

U.D.C. 621.373.51.029.63/4

A Tunnel Diode Oscillator with Wide Tuning Range (0·7–4·9 Gc/s)

By

D. R. PERSSON, M.A.†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July 1965.

Summary: The tunnel diode is coupled weakly to a relatively high Q resonator essentially of the hybrid, or re-entrant coaxial, type. This construction ensures a wide tuning range and a frequency of oscillation that is relatively insensitive to temperature change, supply voltage variations and interchange of diodes. In the basic form of the oscillator, using a 2 mA peak current germanium diode, stable and continuous oscillation is sustained over the full tuning range of 0·7 to 4·9 Gc/s with a power output greater than 1 μW up to 4 Gc/s. Alternative forms of the oscillator still under development give power outputs approaching 1 mW at some cost in tuning range.

1. Introduction

The simplest forms of tunable tunnel diode oscillator rely on modifying the natural self-resonant frequency defined by the junction capacitance and series inductance of a tunnel diode by means of an external inductance or capacitance.

These simple inductively and capacitively tuned circuit configurations are described in the literature and characterized as oscillators operating respectively below or above the self-resonant frequency.[1, 2]

The octave bandwidth oscillator described by Shelton and Frederick[3] would require a more complicated equivalent circuit but the design starting point is clearly the capacitance of the tunnel diode, so that this oscillator is similarly dependent on what might be considered an incidental or stray characteristic of the diode used.

The oscillator design described here was conceived differently as one whose oscillation frequency, in particular, would be that of an associated resonator and relatively independent of the parameters of the tunnel diode itself.

Since a suitably biased tunnel diode presents a negative resistance or conductance to its terminals at all frequencies up to its cut-off frequency it is capable of oscillating at the resonant frequency of, for example, any high-Q series-tuned termination with smaller resistance or high-Q parallel-tuned termination with smaller conductance. To construct a wide range tunnel diode oscillator it is therefore sufficient in principle to offer up to the terminals of a biased tunnel diode a lightly-loaded resonator of the required tuning range. Apart from the broad tuning facility, such a construction could be expected to offer advantages in frequency stability and interchangeability of diodes.

† Formerly with Decca Radar Ltd., Research Laboratory, Walton-on-Thames, Surrey; now with G.E.C. (Electronics) Ltd., Stanmore, Middx..

2. The Resonator

The form of resonator chosen was the hybrid cavity (Fig. 1) which has been described in its application as a wavemeter by Essen.[4]

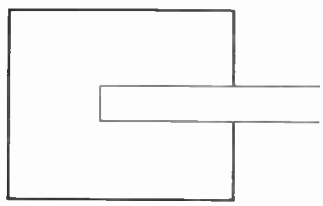


Fig. 1. Hybrid cavity.

At the high frequency end of the tuning range, with the tuning plunger withdrawn, the resonator is a cylindrical cavity with a resonant wavelength, in the E_{010} mode, of 1·3 times the cavity diameter. As the plunger is intruded into the cavity its inductance resonates with its capacitance to the side and end walls of the cavity at frequencies which in theory can be reduced without limit. In practice the very high tuning slope as the plunger almost touches the far end of the cavity restricts the useful tuning range to an octave or two.

In a wavemeter application a detector diode could be coupled to the resonator either capacitively by a probe or inductively by means of a loop as shown in

Fig. 2. Although loop coupling is possibly more usual, capacitive coupling of an axially mounted detector has been employed by the author in a precision 1·5 to 4 Gc/s wavemeter. However in a tunnel diode device the self-inductance of a coupling loop would introduce a stability problem which can be avoided by a capacitive coupling arrangement. Although considered originally, inductive coupling of the tunnel diode was rejected in favour of capacitive coupling.

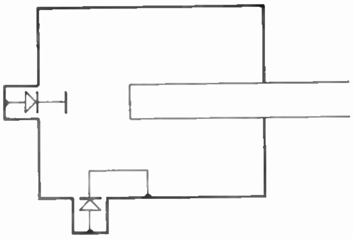


Fig. 2. Hybrid cavity showing alternative capacitively and inductively coupled diodes.

An equivalent circuit of the hybrid cavity with capacitively coupled diode is shown in Fig. 3, where C_1 and C_2 are the capacitances from the tuning plunger to the coupling probe and cavity body respectively and C_0 the capacitance from the probe to the cavity body. To reach low frequencies either C_2 or C_1 and C_0 must become large. These alternative requirements suggest a choice of configurations. The oscillator could be constructed as one in which the diode is series coupled by C_1 to a parallel tuned resonator L, C_2 or as one in which the diode is shunt coupled by C_0 to a series tuned resomator L, C_1.

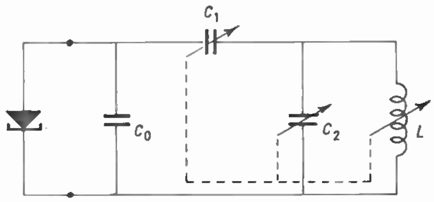


Fig. 3. Equivalent circuit of hybrid cavity with capacitively coupled diode.

Pursuit of the second course led fairly naturally to the construction shown in Fig. 4. Here the probe has become a cylindrical cup extending into the cavity and overlapping the end of the tuning plunger when it is fully inserted. This arrangement allows a smooth transition to the large capacitance required to tune the cavity to the low frequencies. Figure 4 shows also the bias connection and the adjustable output coupling loop with which the cavity is loaded. The diode is held in position mechanically by a clamp which is not shown in detail.

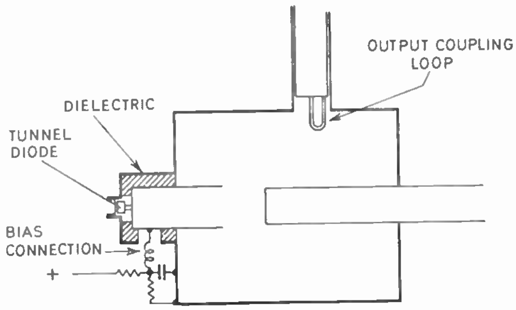


Fig. 4. Construction of wide-range tunnel diode oscillator.

A simplified equivalent circuit is shown in Fig. 5, where R_L represents the external loading and R_d the internal dissipation. With C_0 generally greater than C_1 the main resonant frequency determining parameters are L and C_1, while C_0 determines the degree of coupling of the diode to the cavity. So long as the negative Q of the cavity loaded by the diode alone is smaller in magnitude than the unloaded $Q, \omega_0 L/R_d$, the diode can still oscillate at and stay locked to the resonant frequency of the cavity if this is sufficiently lightly loaded.

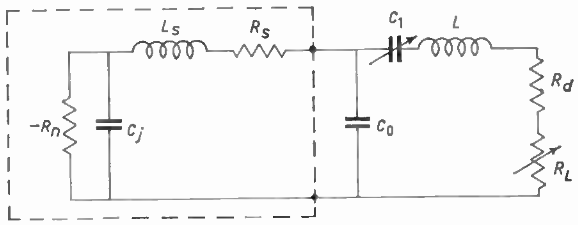


Fig. 5. Simplified equivalent circuit of oscillator.

3. The Diode

In general the choice of tunnel diode for a given oscillator application is governed by the required power output, which is a function of the semiconductor material and of the diode peak current, and the maximum required frequency of operation, which determines the minimum cut-off frequency of the diode. These considerations are subject however to an overriding requirement of stability against spurious modes of oscillation which places constraints on the design and may entail some compromise of the power requirement.

The order of magnitude of the power output can be estimated quite simply. If the r.f. voltage and current at the diode terminals are assumed to be sinusoidal with amplitudes $\frac{1}{2}(V_v - V_p)$ and $\frac{1}{2}(i_p - i_v)$, where V_p, V_v, i_p and i_v are the peak and valley voltages and currents, the expected power output can be taken as half the product or

$$P = \tfrac{1}{8}(V_v - V_p)(i_p - i_v)$$

$$\simeq \tfrac{1}{8}(V_v - V_p)i_p \qquad \ldots\ldots(1)$$

$$\simeq 0{\cdot}04\, i_p \text{ watts} \qquad \ldots\ldots(2)$$

for germanium diodes. In the author's experience the available power output lies between 0·02 i_p and 0·03 i_p. Gallium arsenide diodes, with a larger peak to valley voltage difference, may be expected to give a proportionately higher power output.

The cut-off frequency is defined in terms of the negative resistance, junction capacitance and series resistance of the diode by the relation

$$f_{co} = \frac{1}{2\pi R_n C_j} \cdot \sqrt{(R_n/R_s) - 1} \qquad \ldots\ldots(3)$$

$$\simeq \frac{1}{2\pi C_j \sqrt{(R_n R_s)}} \qquad \ldots\ldots(4)$$

Since the available power output falls rapidly as the frequency of operation approaches the cut-off, and as the effective series resistance of the diode may be increased by the contact resistance and local surface resistance of its mounting, it is generally desirable that the nominal cut-off frequency should exceed the greatest frequency of operation by a factor of two or more.

The actual choice of diode for the first experimental 1–4 Gc/s oscillator was limited by availability to the D4168 series of 2 mA peak current germanium diodes of which the D4168A or D4168B were expected to be the most suitable. A power output of microwatts was quite acceptable for the original broadband signal source application.

4. Stability Considerations

As the coupling between the diode and the principal mode of the cavity resonator is reduced the diode becomes increasingly liable to oscillate preferentially in alternative modes, in particular at its own natural frequency of oscillation or 'self-resonant' frequency. It is therefore desirable that these modes should be fairly heavily damped.

In the case of the self-resonance the damping is provided mainly by the series resistance of the diode so that the effective criterion for stability against oscillation at the self-resonance frequency is simply that this frequency should be greater than the resistive cut-off frequency.

The self-resonance frequency is given by

$$f_0 \simeq \frac{1}{2\pi\sqrt{(L_s C_j)}} \qquad \ldots\ldots(5)$$

where L_s and C_j are the package inductance and junction capacitance of the diode. The criterion for stability is therefore that

$$f_{co} < f_0 \qquad \ldots\ldots(6)$$

i.e.

$$\frac{L_s}{C_j R_n R_s} < 1 \qquad \ldots\ldots(7)$$

from (4) and (5).

This condition is satisfied for example by the 'typical' D4168B diode with $C_j = 1{\cdot}7$ pF, $R_n = 55\,\Omega$, $R_s = 4\,\Omega$, and $L_s = 0{\cdot}25$ nH. (These were the appropriate figures when this development was initiated.)

Other spurious oscillations liable to occur are those associated with the external bias circuitry or with the higher order resonances of the cavity. These oscillations were avoided by careful choice of bias circuit component values in the first case and in the second by a choice of cavity dimensions which ensured that the resonant frequency of the three quarter mode in particular would be close to or higher than the diode cut-off frequency.

5. Performance

In its original form the oscillator gave the power output shown in Fig. 6 (curve A) with a D4168B diode. The output, although feeble at the top end of the tuning range, was maintained throughout the tuning range of 0·7 to 4·9 Gc/s. The high frequency fall in power was due as much to the diode as to losses in the cavity. D4168A diodes with lower cut-off frequencies would not oscillate over the full frequency range. Diodes with much higher cut-off frequencies were unstable, and some D4168B diodes with cut-off frequencies well above the typical values gave spurious outputs in the region of 9 Gc/s.

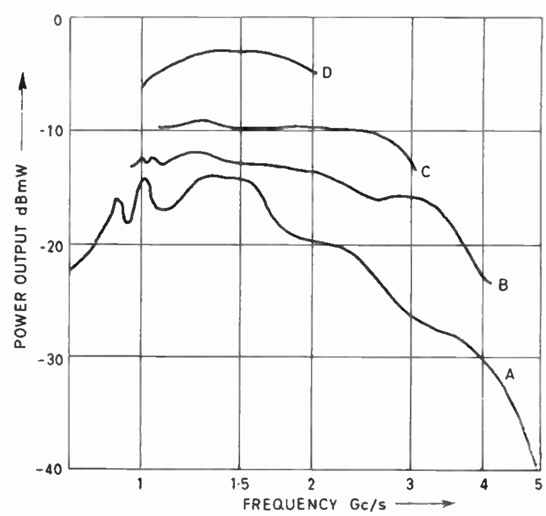


Fig. 6. Power output of original oscillator with D4168B diode (A), modified for D5061A (B), modified experimentally for D5062 and D5064 diodes (C and D).

A revised version of the design gave the improved high frequency performance shown in Fig. 6 (curve B) using a D5061A diode. This is also a 2 mA peak current diode but has a lower inductance package and is rather more closely specified than the D4168 series, so that specification or selection of suitable diodes does not present a very great problem.

Experimental cavities of similar design have been built for higher peak current diodes. Curves C and D in Fig. 6 show the best results achieved so far with 5 and 20 mA peak current germanium diodes. The equivalent gallium arsenide diodes could be expected to give about twice these power levels.

However the higher power oscillators present rather special difficulties. Because of the increased stability problems associated with the higher current diodes and the greater importance anyway of achieving maximum output power the useful bandwidth is reduced and, for 1 mW L- or S-band oscillators, may be limited to about an octave.

The short-term frequency stability of the original form of the oscillator is of the order of a few parts in 10^8. The temperature drift at 2 Gc/s is 200 kc/s or 0·01% per degree C.

6. Conclusions

Application of the design principle of locking a tunnel diode to a high Q resonator leads to an oscillator construction featuring good frequency stability and a wide tuning range approaching three octaves. Power outputs of the order of a milliwatt adequate for laboratory signal sources or receiver local oscillators can probably be achieved over octave bandwidths at L- or S-band.

7. Acknowledgment

A large part of the development and testing of this oscillator was carried out by Mr. R. B. G. Friday.

8. References

1. S. P. Gentile, 'Basic Theory and Application of Tunnel Diodes', p. 190 (Van Nostrand, New York, 1962).
2. R.C.A. 'Tunnel Diodes'. Technical Manual TD30, pp. 67–71 (Radio Corporation of America, 1963).
3. W. L. Shelton and W. L. Frederick, 'Octave tunable tunnel diode oscillators', *Microwave J.*, **5**, pp. 192–5, September 1962.
4. L. Essen, 'Cavity resonator wavemeters. Simple types of wide frequency range', *Wireless Engineer*, **23**, pp. 126–32, May 1946.

Manuscript first received by the Institution on 2nd April 1965 and in final form on 14th May 1965. (Paper No. 1038).

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **March 1966**

March 1966	GBZ 19·6 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 24-hour mean centred on 0300 U.T.	March 1966	GBZ 19·6 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 24-hour mean centred on 0300 U.T.
1	− 301·9	− 301·6	− 1·7	16	− 299·9	− 300·2	− 2·1
2	− 299·0	− 300·7	− 2·0	17	− 298·9	− 300·0	− 2·0
3	− 298·9	− 300·7	− 2·8	18	− 299·3	− 299·7	− 1·7
4	− 299·6	− 300·8	− 2·3	19	− 301·5	− 300·0	− 1·9
5	− 300·8	− 301·1	− 1·8	20	− 300·5	− 300·0	− 1·9
6	− 300·3	− 301·3	− 1·3	21	− 299·0	− 299·1	0
7	− 299·2	− 300·3	− 1·3	22	− 299·5	− 300·2	− 2·1
8	− 298·9	− 299·9	− 1·7	23	− 298·1	− 299·7	− 2·5
9	− 299·7	− 300·4	− 1·5	24	—	− 299·6	− 2·6
10	− 298·0	− 301·0	− 2·2	25	− 300·0	− 298·7	− 1·8
11	− 299·5	− 300·1	− 2·3	26	− 298·0	− 300·1	− 0·9
12	− 299·1	− 299·8	− 1·8	27	− 297·8	− 300·6	− 0·1
13	− 300·0	− 299·9	− 1·6	28	− 301·4	− 300·7	− 0·1
14	− 301·1	− 300·5	− 1·6	29	− 299·9	− 300·6	− 0·2
15	− 300·8	—	− 1·8	30	− 300·6	− 300·9	− 0·1
				31	− 299·9	− 301·5	− 0·6

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

Note: GBR Rugby has been replaced temporarily by GBZ Criggion.

U.D.C. 621.382.23

Gallium Arsenide Varactor Diodes

By

C. A. P. FOXELL, B.Sc.†

AND

K. WILSON, B.Sc.(Eng.)†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July, 1965.

Summary: From consideration of the factors influencing the design of varactor diodes for use in high-frequency parametric amplifiers and multipliers it appears that GaAs is the most suitable choice of semiconductor. The high carrier mobility of this material over a wide range of temperatures should enable high cut-off frequency varactor diodes to be fabricated which will operate at the low temperatures required for ultra-low-noise parametric amplification. In addition, the high cut-off frequency should provide efficient harmonic generation at the higher microwave frequencies.

An appraisal of the performance of an initial GaAs diode indicated that low frequency measurements were unreliable, and this has led to the development of a microwave measurement technique which has proved to be an extremely useful method of evaluating the junction parameters.

As a result of these measurements new device structures have been evolved which reduce the microwave losses and simplify the fabrication processes. Three forms of diode have been developed for use at C, X and Q-band having cut-off frequencies of 150, 250 and 700 Gc/s (at zero bias) respectively, and the construction and performance of these varactors are described.

1. Introduction

The variable capacitance characteristics of a semiconductor diode are due to the mobile charge carriers moving away from the neighbourhood of the p–n junction to form a space charge or depletion layer. The width of this depletion layer is determined by the bias applied to the junction and therefore the capacitance can be varied between the extreme points of forward and reverse conduction. Such a device exhibiting a voltage dependent junction capacitance is known as a varactor diode and these have recently found particular application in the fields of low-noise parametric amplification and harmonic frequency generation. The conventional equivalent circuit of a varactor diode is shown in Fig. 1 in which C_j the junction capacitance charges through a series resistance R_s whilst the encapsulation provides a stray capacitance C_s and inductance L_s.

Although suitable varactor diodes for low frequency operation have been available for some time, the lack of high performance devices has tended to limit progress on applying these techniques at microwave frequencies. The present paper reviews the basic requirements of a varactor diode for use at microwave frequencies and considers the reasons for employing gallium arsenide as the semiconducting material to obtain optimum performance. Apart from the choice of semiconductor, the characteristics of a varactor diode also depend to a considerable extent upon the physical structure of the device, and the development of a microwave measurement technique for evaluating the true diode parameters has enabled considerable improvements to be made in the design of this type of diode. On this basis a range of diodes have been investigated for use from 1 to 40 Gc/s, and the construction and performance of these will be described.

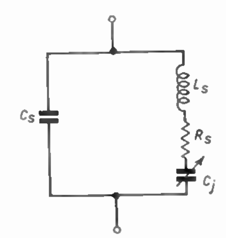


Fig. 1. Conventional varactor diode equivalent circuit.

2. Diodes for Parametric Amplifiers and Harmonic Generators

It is possible to calculate the idling frequency at which a parametric amplifier noise figure has its minimum value.[1] This is given by:

$$f_2 = -f_1 + (f_1^2 + \gamma^2 f_c^2)^{0 \cdot 5} \qquad \ldots\ldots(1)$$

† Associated Semiconductor Manufacturers Limited, Wembley Laboratories, Wembley, Middlesex.

where f_1 = signal frequency

f_2 = optimum idling frequency

γ = capacitance (C) variation coefficient

$$\frac{C_{max}-C_{min}}{2(C_{max}+C_{min})}$$

f_c = diode cut-off frequency

The minimum noise figure theoretically predicted is then

$$F_{min} = 1+\frac{2f_1}{\gamma f_c} \qquad \ldots\ldots(2)$$

$$\left(\text{assuming } \frac{\gamma f_c}{f_1} \gg 1\right)$$

A suitable figure of merit for a varactor diode in this application is therefore,

$$M = \gamma f_c \qquad \ldots\ldots(3)$$

and the relation between the estimated amplifier noise figure and this figure of merit is given in Fig. 2 for a range of signal frequencies. The amplifier noise figure from eqn. (2) can also be expressed in terms of noise temperature and becomes,

$$T_{amp} = \frac{2f_1}{M}T \qquad \ldots\ldots(4)$$

where T_{amp} = amplifier noise temperature

T = diode temperature

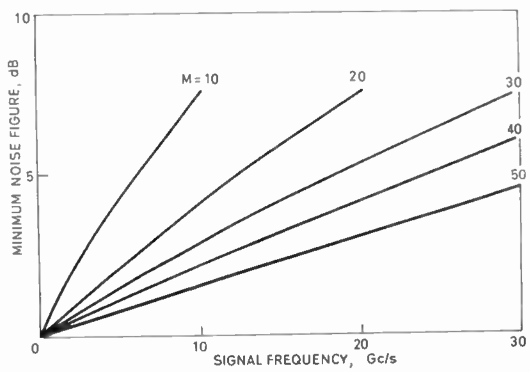


Fig. 2. Relation between diode figure of merit and parametric amplifier performance.

From these last two equations it becomes apparent that for optimum amplifier performance a diode should combine a high cut-off frequency with a marked capacitance variation and that these features should be maintained at low temperature in order to achieve low noise figures.

In the case of harmonic generation, the two diode parameters of particular importance are the power conversion efficiency and the power handling capacity. Penfield and Rafuse[2] have demonstrated how the conversion efficiency, η, is approximately related to the device characteristics as,

$$\eta = 1-D\frac{f_{in}}{f_c} \qquad \ldots\ldots(5)$$

where f_{in} = fundamental frequency

D = constant depending on order of harmonic desired

which indicates that, as for parametric amplifiers, a high cut-off frequency is desirable. The power handling capacity of the diode is governed by the junction capacitance, permissible voltage swing, and the power dissipation rating. As long as the dissipation rating is not exceeded, the maximum permissible input power P_{in} is given by,

$$P_{in} = Af_{in}C_{min}(\phi+V_B)^2 \qquad \ldots\ldots(6)$$

where ϕ = junction contact potential.

The constant A will depend primarily on the order of harmonic selected and the frequency of operation. However, another factor which complicates this analysis is that if the lifetime of the minority carriers in the semiconductor is sufficiently long the diode may be driven into the forward direction to obtain a significantly large diffusion capacitance which will enhance performance. Although this effect has been utilized at lower frequencies[3] its importance in the microwave region has not been determined.

As the junction capacitance, C_j, is restricted by cut-off frequency and impedance level considerations it is necessary to counterbalance this with a high breakdown voltage, V_B for large input power levels.

The thermal rating is determined by the maximum junction temperature rise that the diode structure can tolerate, T_j, and the thermal resistance θ. Then the maximum dissipation within the device becomes,

$$P_{diss} = \frac{T_j}{\theta} \qquad \ldots\ldots(7)$$

Consideration of these diode requirements for amplifiers and multipliers indicates that there are several parameters which are common for both applications. The outstanding features that such an ideal diode should possess are a high cut-off frequency combined with a large capacitance-voltage sensitivity. Whilst the relative importance of the other characteristics such as high breakdown voltage, low temperature operation, and good thermal dissipation will depend on the particular application of the diode, it should be the aim to bring these features together in one device.

3. Diode Design

The design of microwave varactor diodes can be resolved into the semiconductor material, the junction element, and the device encapsulation.

Varactor diodes have been made from several kinds of semiconductor. It can be shown that the cut-off frequency which is defined by,

$$f_{c(V)} = \frac{1}{2\pi R_s C_{j(V)}} \qquad \text{......(8)}$$

can be expressed in terms of the properties of the material as follows

$$f_{c(V)} \propto \mu . \varepsilon^{-\frac{1}{2}} \qquad \text{......(9)}$$

Thus the perfect semiconductor should possess a low dielectric constant and a high minority carrier mobility. On this basis indium antimonide (μ = 100 000 cm^2/Vs, ε = 17) would appear outstanding, and varactors fabricated with this material have been reported,[4] with a cut-off frequency of 400 Gc/s which provided extremely low noise amplification. However, as this material has an energy gap of only 0·18 eV it can only be used below about −196°C and this considerably restricts the usefulness of this type of diode. Of the materials which can operate over a wider range of temperatures gallium arsenide (μ = 4 500 cm^2/Vs, ε = 11·1) and silicon (μ = 1 000 cm^2/Vs, ε = 11·8) appear the most promising. The energy gap of silicon (1·08 eV) allows operation up to 175°C but at lower temperatures the conductivity decreases due to the high impurity ionization energies. Recently some improvement has been described[5] by using compensated material, but cut-off frequency still decreases by 50% at −269°C as compared to 25°C. Therefore from the viewpoint of cut-off frequency and ability to operate over the temperature range of −269°C to 225°C gallium arsenide seems to be the most suitable material for parametric amplifier use. Whilst these are also essential characteristics for harmonic generator applications progress in this field has been retarded by the difficulty of achieving high breakdown voltages in gallium arsenide. This has been a result of the technological problems associated with producing material of reasonable purity. However, the introduction of epitaxial growth processes has opened the way to better quality gallium arsenide and breakdown voltages of greater than 100 V are now feasible.

In the previous section it was shown that the electrical performance of a varactor diode is defined by cut-off frequency, capacitance variation and breakdown voltage. Therefore, the design of the junction element depends on producing the optimum relationship between the series resistance and capacitance for a given breakdown voltage. Breakdown voltage V_B is a function of the gallium arsenide resistivity ρ_n and with the types of junction described in this paper it has been found that,

$$V_B = 250\rho_n^{0\cdot6} \qquad \text{......(10)}$$

over the range of 5 to 100 V. For a parametric amplifier only a low breakdown voltage is required (i.e. about 6 V), but it is always advantageous to use the lowest possible resistivity in order to minimize the series resistance.

The resistance of the diode arises not only from the resistive components of the junction and bulk semiconductor, which can be calculated from the dimensions and resistivity,[6] but also from contact resistances and skin effects at microwave frequencies. Although these are the elements that in practice limit gallium arsenide varactor diode performance they are not fully understood.

The junction capacitance is also dependent on resistivity as

$$C_{j(V)} = A_j \left[\frac{2\rho\mu(V-\phi)}{\varepsilon}\right]^{-n} \qquad \text{......(11)}$$

Ideally the junction capacitance should be larger than any fixed stray capacitance from the encapsulation, and yet less than about 1 pF for a suitable microwave impedance. As a low capacitance is required for a high cut-off frequency this implies a high resistivity and a small junction area A_j. Unfortunately this is in opposition to the low resistivity and large junction area required to reduce the series resistance, and therefore some compromise is necessary. In practice the optimum $R_s C_j$ product is obtained with a junction capacitance in the region of 0·1 pF.

The value of n is derived from the impurity distribution at the junction. For a graded junction this is about $\frac{1}{3}$ whereas for an abrupt type n is $\frac{1}{2}$. To enhance γ, the capacitance variation coefficient, it is desirable to approximate as closely as possible to an abrupt junction.

Basically, the encapsulation can limit the performance of a varactor diode by having: (*a*) a large stray capacitance in relation to the junction capacitance, (*b*) an excessive series inductance giving low series and parallel resonant frequencies and, (*c*) poor thermal properties. Most varactor cases consist of two metal connections carrying the diode element separated by a ceramic insulator. In general, reducing the size of the encapsulation by bringing the metal parts closer together will give a lower series inductance, but due to limitations on the minimum size of ceramic insulator that can be fabricated this will tend to increase the stray capacitance and therefore some compromise must be made between the stray capacitance and inductance. Apart from the inductance of the metal case components there is a contribution from the internal connections to the diode element and these will be discussed in the sections dealing with construction of actual devices.

Due to the practical limitations on the reduction of size of the case it becomes increasingly difficult to attain series resonant frequencies above about 30 Gc/s. Mounting the diode junction element directly in a section of wave guide then becomes a more attractive solution and a diode of this type will also be mentioned.

The thermal aspects of the case design are those of good power dissipation from a low thermal resistance (eqn. (7)), and the ability to function over a wide range of temperatures, probably from $-269°C$ to $225°C$. Ceramic-metal encapsulations are ideally suited to these conditions and the main considerations are to employ metal parts of high thermal conductivity yet matched in expansion to the ceramic.

4. Diode Evaluation

Measurement of the various diode parameters is necessary for the correct assessment of a device and also to provide a firm basis for further research and development. Some characteristics such as the breakdown voltage can be easily determined by simple d.c. measurements, but other parameters are frequency dependent and it is therefore essential to evaluate microwave varactor diodes at frequencies about which they will be employed.

A most satisfactory microwave transmission measurement technique has been established and is fully described in another paper.[7] Transmission loss past the diode is measured in the region of the diode series resonant frequency, which in the present instance is usually in X or Q-band. The diodes are mounted in reduced height waveguide with tapers matching to full height waveguides on either side. The equivalent circuit is then assumed to be that shown in Fig. 1, shunting the transmission line. In the region of the series resonant frequency the impedance of the series branch is very low, so the shunting effect of the case capacitance and the detector may be neglected. The equivalent circuit then becomes that of a series inductance-capacitance-resistance circuit shunting the transmission line. Three measurements, of transmission loss at resonance, Q factor, and series resonant frequency enable the three unknowns to be determined. In addition, the variation of the resonant frequency with bias enables the capacitance variation with bias to be determined.

Two salient points emerged from the measurement of some early diodes by this method. The first was an increase in the series resistance of up to 5 ohms over the d.c. value of 1 ohm, and the second feature was that the values obtained for the junction capacitance were about 0·1 pF higher than measurements made on a 1 Mc/s bridge. In view of the fact that the values of series inductance agreed well with the inductance of dummy metal diodes it is thought that the microwave values of junction capacitance and series resistance do correctly describe the diode microwave performance. Further work suggests that this increase in junction capacitance is due to a distributed case capacitance, part of which appears across the junction, and therefore the usual equivalent circuit should be modified to that shown in Fig. 3. As mentioned previously, the increase in series resistance arises from skin effects at microwave frequencies. The determination by the microwave measurements of these significant changes in junction capacitance and series resistance, which could easily reduce the cut-off frequency of a diode to 20% of the low frequency value, had an important bearing on the subsequent development of the varactor diodes. With this measurement technique it has been possible to evaluate various designs and minimize the degradation in performance at microwave frequencies.

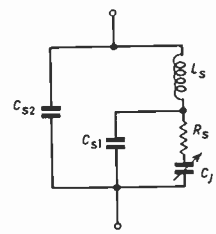


Fig. 3. Modified varactor diode equivalent circuit.

5. Diode Development

The rest of the paper describes the construction and performance of three different gallium arsenide varactor diodes which mark the present state of development. These are primarily intended for use in S to C, X, and Q-band parametric amplifiers, but should also prove useful as harmonic generators over the same frequency range. The minimum basic requirements for these diodes, based on the factors outlined earlier in Section 1, are shown in Table 1, and these parameters should be maintained at low temperatures for cooled amplifiers.

6. S to C-Band Varactor Diode

6.1. *Diode Fabrication*

A small area diffused junction diode is a convenient structure for obtaining optimum characteristics where the desired junction capacitance is above about 0·1 pF. The smaller areas required for capacitances below this value involve contact problems and a point-contact junction becomes more attractive. However in this

Table 1

Minimum varactor diode requirements

Amplifier frequency	S to C-band	X-band	Q-band
Type No.	VX3368 (Mullard CAY10)	VX6508	Experimental
$M = \gamma f_{c(0v)}$†	⩾ 20 Gc/s	⩾ 30 Gc/s	⩾ 50 Gc/s
Series resonant frequency	⩾ 8 Gc/s	⩾ 15 Gc/s	⩾ 30 Gc/s

† $\gamma = \dfrac{C_{max}-C_{min}}{2(C_{max}+C_{min})}$ where $C_{min} = C_j$ at $V_R = 1$ V
$C_{max} = C_j$ at $I_f = 1\ \mu$A
and V_R and I_f are reverse voltage and forward current respectively.

instance a capacitance of the order of 0·5 pF is desirable. Diffusion is not only a controllable process, but also a flexible technique which can be applied with epitaxial and planar structures to give further improvements in the electrical characteristics. In addition, diffusion in gallium arsenide has another significant advantage. Unlike the majority of diffusants in other semiconductors which produce an error function profile resulting in a graded junction, the diffusion of zinc in gallium arsenide is anomalous and approaches the heavily doped abrupt profile given by an alloyed junction. Therefore by employing a zinc diffusion to obtain a junction in n-type gallium arsenide a greater capacitance-voltage sensitivity is achieved.

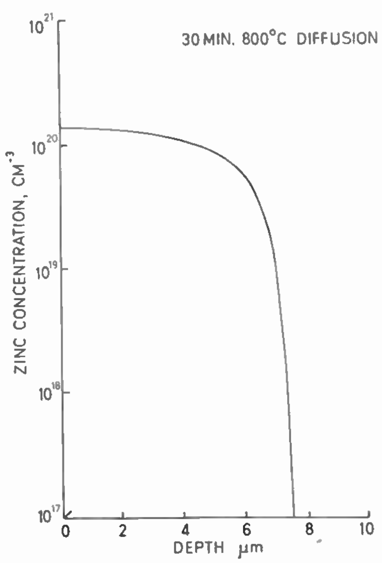


Fig. 4. Typical zinc diffusion profile in gallium arsenide.

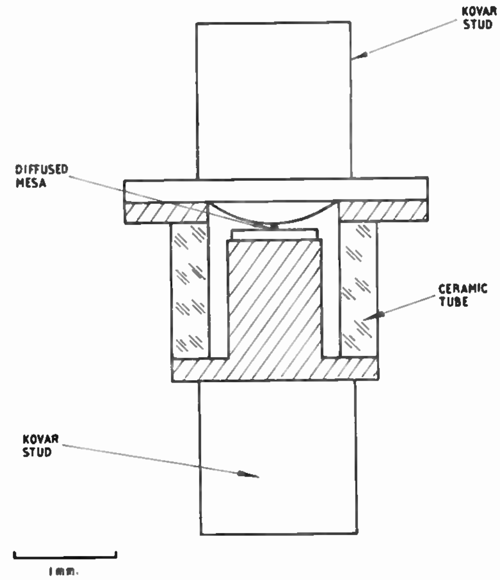


Fig. 5. VX3368 (CAY10) C-band varactor diode.

By using n-type gallium arsenide in the form of a thin slice of the lowest possible resistivity compatible with the required breakdown voltage of 6 V (see eqn. (10)) the bulk resistance of the device is minimized. After chemically polishing the slice to obtain a flat surface, the zinc is diffused in under an excess pressure of arsenic to prevent disassociation of the gallium arsenide. By carrying out the diffusion at a high temperature for a short time an extremely heavily-doped abrupt p-type layer is obtained as shown in Fig. 4. In gallium arsenide the gallium and arsenic occupy alternate sites in the crystal lattice and this asymmetry is the cause of many of the special characteristics of the material. In consequence the junction on the gallium rich face is removed to retain the flatter junction on the arsenic rich side. The importance of obtaining a flat junction is that the thickness of the p-type layer can be minimized to limit the resistance R_p, of this part of the junction.

This slice containing a p–n junction is now diced into wafers about 750 microns square and mounted by alloying into the pill case as shown in Figs. 5 and 6.

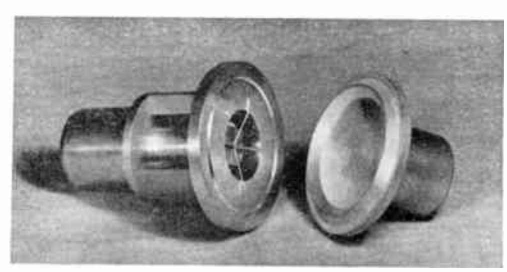


Fig. 6. Construction of the VX3368 (CAY10) C-band varactor diode.

The importance of this mounting process was revealed by measurements on early diodes which indicated that the contact resistance R_c could amount to several ohms. However, by suitable choice of alloying materials this has been reduced to about 0·5 ohms. A metal contact is now attached to the p-type layer and connected to the top flange of the pill case by means of thermo-compression bonded gold wires. As full connections are now made to the diode it is possible to monitor the junction characteristics whilst the junction size is reduced by etching until the capacitance is about 0·5 pF. The diode is completed by resistance welding a top cap to hermetically seal the pill case.

6.2. *Encapsulation*

The combination of a ceramic-metal case with a diode containing all bonded or alloyed connections makes a very rugged device, and by matching the thermal expansion of the various components operation is possible down to liquid helium temperatures.

With a case of given outline a compromise must be made between the inductance and capacitance. As an inductance, the case consists of a pillar on which is mounted the junction with a wire connection to the top cap, and as the overall length of the case is fixed, any change in pillar height is counteracted by a corresponding change in wire length. The relative inductances of these two components were determined at 10 Gc/s and the conclusion was drawn that the maximum pillar height that could be accommodated in the case would provide the lowest inductance. The material has little effect, and whilst an increase in pillar diameter would be worthwhile this is restricted by the overall outline of the case. Although enlarging the junction connection is also beneficial this leads to increasing difficulties in fabricating the device.

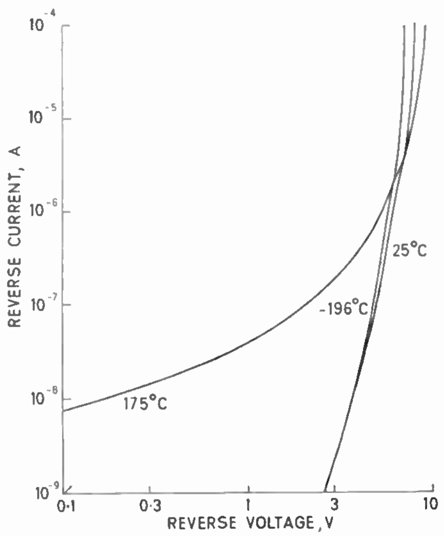


Fig. 7. Reverse characteristics of a diffused gallium arsenide varactor diode.

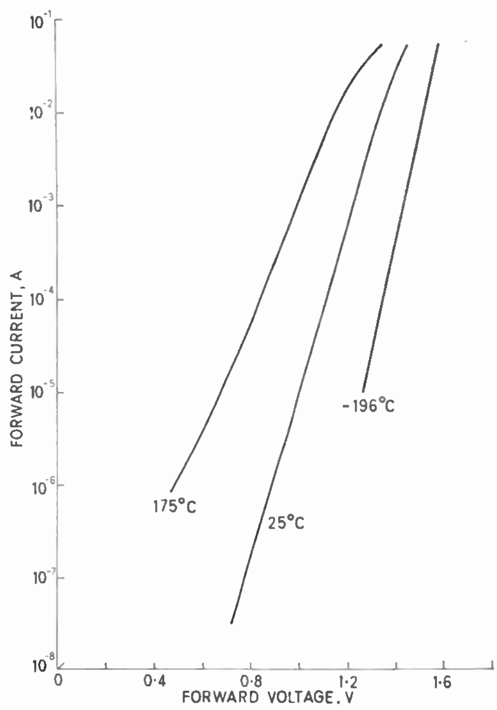


Fig. 8. Forward characteristics of diffused gallium arsenide varactor diode.

From the measurements described in Section 4 it was found that there was a stray capacitance across the junction (C_{s1}) and a stray capacitance due to the encapsulation (C_{s2}). The direct capacitance between the wafer and the top cap is negligible and C_{s2} is the capacitance from the sides of the pillar via the ceramic insulator tube to the top cap. For minimum stray capacitances this tube should have a small cross-sectional area and maximum height, but outside diameter and height are restricted by the use of a standard outline, and the wall thickness is limited by porosity considerations to about 375 microns. By employing high aluminium ceramic a rugged seal is obtained, but if a technique for metallizing quartz becomes available the low dielectric constant of this material would drastically reduce the stray capacitance.

6.3. *Performance*

The typical reverse characteristics of the diode are shown in Fig. 7 for ambient temperatures of 175°C to −196°C. As a result of the large energy gap of gallium arsenide the reverse leakage currents are extremely low being of the order of 10^{-8} A at room temperature, and only becoming significant near the maximum operating temperatures of 175°C. The corresponding forward characteristics for these temperatures are given in Fig. 8 and it will be noted that these changes follow from an increase in the diode contact potential ϕ of about 2 mV for every 1°C fall in temperature. This effect in conjunction with the high value of contact potential, resulting from the energy gap, enhances the capacitance variation coefficient at low temperatures. These diodes withstand the thermal shocks involved in rapid changes over this temperature range and some devices have been stored at −269°C. Although no precise measurements were made at this temperature the devices still exhibited good diode characteristics.

The capacitance law of the diodes, when measured on a low frequency bridge, followed the relationship,

$$C_{j(V)} = C_{j(0V)}(1-V)^{-0\cdot41} \qquad \ldots\ldots(12)$$

with a typical value of $C_{j(0V)}$ of 0·4 pF. The decrease observed in junction capacitance with temperature correlated with the measured variation in the contact potential.

The d.c. series resistance of the diode can be estimated by extrapolating the forward characteristic, as about 2·0 ohms. The complete low frequency characteristics are summarized in Table 2, and it is interesting to calculate that the cut-off frequency at −6 V is about 442 Gc/s on the basis of these measurements.

Table 2

Typical characteristics of the VX3368 (CAY10) varactor diode

Low frequency characteristics (25°C)	
Breakdown voltage, at $I_R = 1\ \mu A$	6·8 V
Reverse current, at $V_R = 4$ V	20 nA
Forward voltage, at $I_f = 1\ \mu A$	0·88 V
Forward voltage, at $I_f = 50$ mA	1·38 V
Series resistance, R_s	2·10 ohms
Junction capacitance, at $V_R = 6$ V $C_{j(6V)}$	0·18 pF
Stray capacitance, $C_s = C_{s1} + C_{s2}$	0·23 pF
Thermal resistance, θ	500 deg C/W
Microwave characteristics (25°C)	
Series resistance, R_s	2·8 Ω
Junction capacitance, at $V_R = 6$ V $C_{j(6V)}$	0·24 pF
$C_{j(0V)}/C_{j(6V)}$	1·57
Capacitance variation coefficient	0·16
Series inductance, L_s	640 pH
Series resonant frequency, at $V_R = 0$ V, f_{res}	10·2 Gc/s
Q, at f_{res}	14·7
Cut-off frequency, at $V_R = 0$ V, $f_{c(0V)}$	150 Gc/s
Cut-off frequency, at $V_R = 6$ V, $f_{c(6V)}$	235 Gc/s

However, as previously stated, at microwave frequencies there is a modification to the simple equivalent circuit which considerably reduces this figure. The diodes were evaluated by the transmission measurement at the diode series resonant frequency in X-band. The value of microwave series resistance then increased to about 2·8 ohms as a result of skin effects, and the junction capacitances by about 0·1 pF due to the distributed case capacitance. From these measurements it would appear that the correct cut-off frequencies are typically 150 Gc/s at zero bias and 235 Gc/s at −6 V.

The other microwave parameters are given in Table 2 and indicate that a zero bias figure of merit of about 23·5 Gc/s is obtained. From Section 1 such a value would predict a noise figure of about 2 dB at C-band and performance of this order has recently been achieved with a degenerate amplifier.[8]

7. X-Band Varactor Diode

7.1. *Fabrication*

To achieve the performance required for operation at X-band, improvements are necessary in both the diode elements and encapsulation as compared with the previous diode. The possibility of attaining better junction characteristics lies with either increasing the capacitance-voltage sensitivity or the cut-off frequency. Although Chang[9] has described methods of making 'hyper-abrupt' and 'pagoda' structures which enhance the capacitance variation, these usually incur some increase in series resistance and so degrade the overall performance. Alternatively there appear to be several ways to reduce the diode series resistance and thereby improve the cut-off frequency. As shown in Section 4 this resistance arises from the several parts of the diode. In the course of the development of the VX3368 (CAY10) the very high contact resistance of several ohms was drastically reduced giving the approximate resistances shown in Table 3.

Table 3

Estimated components of series resistance at X-band

	VX3368	VX6508
Diffused p-type layer, R_P	0·1 ohms	0·1 ohms
n-type material, R_B	1·4 ohms	0·4 ohms
Contact resistance, R_C	0·5 ohms	0·5 ohms
Increase in R_B due to skin effects R_M	0·8 ohms	0·7 ohms
Total, R_s	2·8 ohms	1·7 ohms

Although the contact resistance could probably be improved still further the major limitations were those of bulk resistance and microwave skin effect. Recently, with the introduction of new methods of producing gallium arsenide it has become feasible to reduce the resistance of the bulk n-type layer. By epitaxial deposition techniques it is possible to grow a thin high-resistivity layer of n-type gallium arsenide on a heavily doped substrate. With material of this type the junction can be formed in the thin high-resistivity layer and the breakdown voltage and other parameters associated with that doping level retained, whilst the substrates being of low resistivity reduces the bulk resistance of the device.

Epitaxial gallium arsenide of this type has been used to fabricate diodes to meet the requirements outlined in Section 5. To obtain the full benefit of this structure it is essential to have a very thin high-resistivity layer as shown in Fig. 9. With such a layer of about 10 microns thickness the diffusion of the p-type junction becomes extremely critical, and involves the control of junction depth to within 1 micron. Apart from this problem of accurate diffusion, the general construction with the steps of mounting, bonding, and etching follows that of the C-band diode. However, the microwave measurements show that the epitaxial material reduces the series resistance by about 1 ohm as indicated in Table 3.

7.2. *Encapsulation*

The series inductance of the case used for the C-band diode is about 640 pH giving a series resonant frequency of about 10 Gc/s with a junction capacitance of 0·4 pF, and it would therefore limit operation at X-band. By employing a very much smaller micropill case, as shown in Figs. 10 and 11, a significant increase in resonant frequency can be achieved. As previously mentioned, the restrictions on fabricating the ceramic insulator tube preclude any reduction in wall thickness but the cross-sectional area can be brought down to the limit set by the space required to house the junction element. In this manner a stray

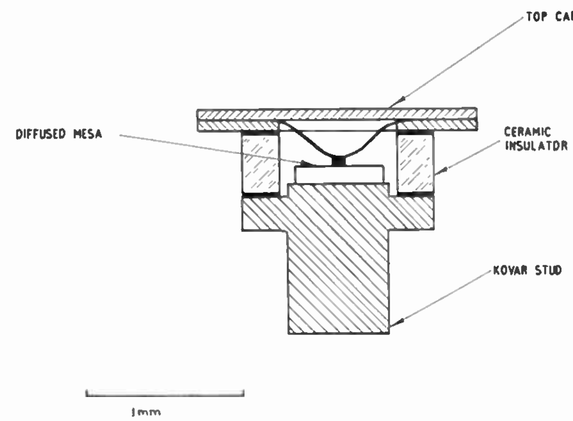


Fig. 10. Structure of VX6508 X-band varactor diode.

capacitance of about 0·35 pF is obtained with a high-alumina ceramic. Although this parasitic capacitance is still of the same order as the VX3368 (CAY10) diode, the decrease in spacing between the metal end connections has a pronounced effect on lowering the series inductance from about 640 pH to about 100 pH. Microwave transmission measurements on this type of diode show that the series resonant frequency has been increased to about 30 Gc/s, thus easily allowing operation at X-band.

7.3. *Performance*

The reverse characteristics and capacitance variation are controlled by the doping of the n-type epitaxial layer and are therefore similar to those of the non-epitaxial diode. However, due to the thinness of this layer and the heavy doping of the substrate the forward voltage drop is appreciably lower and this

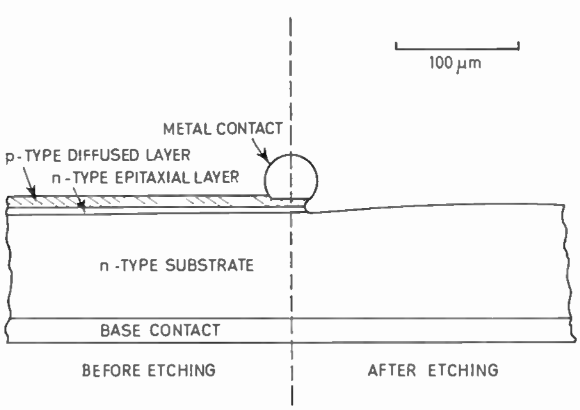


Fig. 9. Diffused epitaxial diode structure.

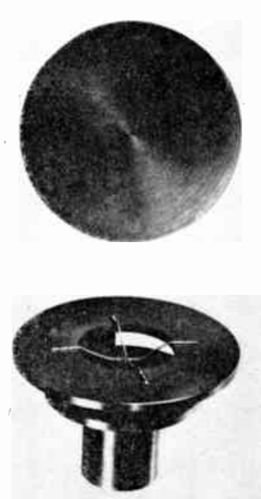


Fig. 11. Construction of the VX6508 X-band varactor diode.

reduces the typical d.c. series resistance to about 1·1 ohms. The microwave transmission measurements indicate that this usually increased to about 1·7 ohms thus giving a cut-off frequency of 250 Gc/s at zero bias with a junction capacitance of 0·3 pF. This is a typical value and cut-off frequencies at zero bias of over 350 Gc/s have been obtained with these diodes. The details of the static and microwave characteristics are listed in Table 4, and from these parameters it will be seen that figures of merit of 30 to 40 have been achieved. As this is combined with an encapsulation resulting in an overall series resonant frequency of 30 Gc/s it should provide an encouraging performance in X-band amplifiers. From Section 1 this figure of merit predicts a noise figure of about 2·5 dB and noise figures of the order of 2·8 dB have recently been obtained with this diode.[10] With a similar varactor diode Sadler and Wells[11] have reported noise figures of 3 dB at room temperature, 1·3 dB (100°K) at liquid nitrogen temperature, and 0·4 dB (17°K) at liquid helium temperature.

Table 4

Typical characteristics of the VX6508 varactor diode

Low frequency characteristics (25°C)	
Breakdown voltage, at $I_R = 1\ \mu A$	5 V
Forward voltage, at $I_f = 1\ \mu A$	0·82 V
Forward voltage, at $I_f = 50$ mA	1·36 V
Series resistance, R_s	1·1 ohms
Junction capacitance, at $V_R = 6$ V, $C_{j(6V)}$	0·17 pF
Stray capacitance, $C_s = C_{s1} + C_{s2}$	0·35 pF
Microwave characteristics (25°C)	
Series resistance, R_s	1·7 ohms
Junction capacitance, at $V_R = 6$ V $C_{j(6V)}$	0·245 pF
$C_{j(0V)}/C_{j(6V)}$	1·53
Capacitance variation coefficient	0·127
Series inductance	120 pH
Cut-off frequency, at $V_R = 0$ V; $f_{c(0V)}$	250 Gc/s
Resonant frequency at $V_R = 0$ V	29 Gc/s

The use of a layer of thin high resistivity gallium arsenide to provide a high breakdown voltage, without incurring the usual degradation in series resistance, has an obvious attraction for harmonic generator applications. For example, by employing such a layer in the VX6508 diode structure it has been found possible to combine a cut-off frequency (at zero bias) of 295 Gc/s with a breakdown voltage in excess of 40 V. It has been reported[12] that with a diode of similar characteristics in a tripler multiplier, outputs of 100 mW have been attained at 34 Gc/s with an efficiency of 50%.

8. Q-Band Varactor Diode

8.1. *Fabrication*

As in the comparison between the C and X-band diodes it is necessary to attain a still higher performance from the junction element and encapsulation for operation at Q-band.

Although the recent experimental work has shown that series resonant frequencies approaching Q-band are feasible when the series inductance is minimized, at these frequencies there is a need for a reduction in the case stray capacitance. This is a function of both the size of the insulator, which has also been reduced to the practical limit, and the dielectric constant of the insulator. Whilst other materials such as beryllia and quartz have lower dielectric constants than the high-alumina ceramic employed in the micropill case, these invariably involve some fabrication problem such as metallizing. Therefore progress with a discrete encapsulation for higher frequency applications awaits advances in ceramic technology. A way of overcoming this difficulty is to insert the unencapsulated junction element directly into the waveguide and this has been the procedure adopted for this experimental work at Q-band.

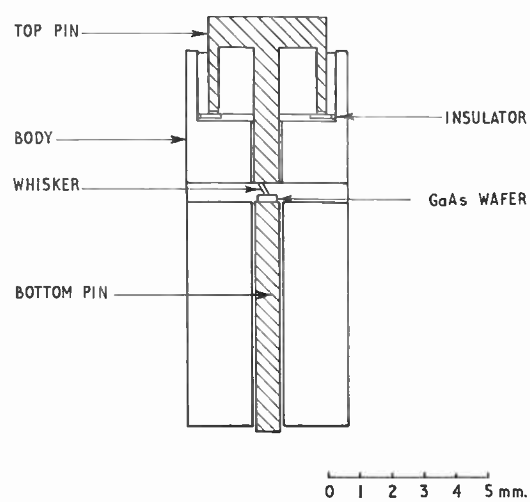


Fig. 12. Experimental Q-band varactor diode.

For optimum performance at these higher frequencies the junction capacitance must be lower, and the dual problems of a smaller junction capacitance and fabricating a diode directly in Q-band waveguide are best solved by a point-contact junction technique. The arrangement adopted is shown in Fig. 12. The body of the diode consists of a section of reduced height Q-band waveguide into which can be inserted two pins, one pin carrying the wafer of semiconductor, and the other the whisker pin which also forms the choke system. The basic method of making the junction is to advance the whisker until contact is made with the wafer and then form a junction by applying low voltage pulses. In this manner the p-type impurity from the wire, usually copper or zinc, diffuses into n-type gallium arsenide to produce a shallow junction of small area.

8.2. *Performance*

This fabricating technique has been combined with the microwave transmission measurement so that it is possible to monitor the varactor characteristics continuously whilst the junction is formed to optimum performance. By inserting the diode waveguide mount into a Q-band transmission line and using a swept bias the microwave transmission characteristics can be displayed on an oscilloscope and compared with the usual current voltage parameters as shown in Fig. 13. The procedure followed is to set the frequency at the desired zero bias resonant frequency, in this instance about 35 Gc/s, and then advance the whisker until contact is made. This is indicated by the d.c. characteristics and resonance at a positive bias. The area of the junction is then increased by electrical forming until the resonance occurs at zero bias. The complete diode performance can then be obtained from the transmission and d.c. characteristic displays. This technique enables the diode characteristics to be controlled very closely and has proved to be a most useful method of investigating variations in the junction structure.

To date only a few experimental diodes have been assembled usually by forming a 37 microns diameter phosphor bronze wire on to 2–5 mΩ-cm n-type gallium arsenide, and some electrical characteristics are given in Table 5.

Table 5

Experimental Q-band varactor diode characteristics

Breakdown voltage, at $I_R = 1\ \mu$A	5 V
Forward voltage, at $I_f = 1\ \mu$A	0·5 V
Characteristics at Q-band	
Junction capacitance, at $V_R = 0$ V, $C_{j(0V)}$	0·065 pF
Series resistance, R_s	3·5 ohms
Cut-off frequency, at $V_R = 0$ V	700 Gc/s
Resonant frequency, at $V_R = 0$ V	35·5 Gc/s
Series inductance, L_s	310 pH
Capacitance variation coefficient, y	0.07

These point-contact gallium arsenide diodes behave in a similar manner to the diffused varactors although most of the diodes at present exhibit some excess leakage current. This tends to depress the value of

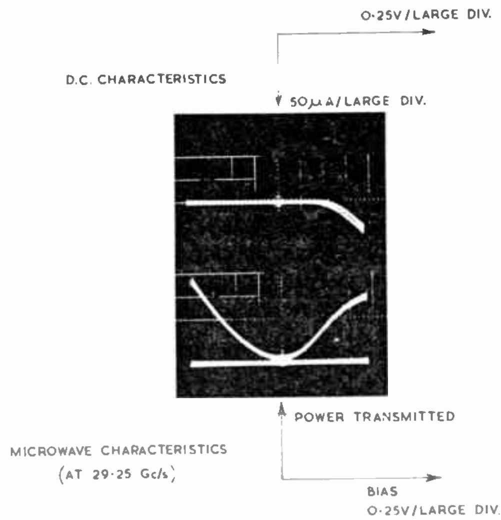


Fig. 13. D.c. and microwave characteristics of a Q-band varactor diode.

capacitance variation coefficient as defined by a level of 1 microampere, but this is not a fundamental limitation and some diodes have been assembled with lower leakage currents. Even so, with the point-contact diode, it seems feasible to obtain high cut-off frequencies, of the order of 700 Gc/s at zero bias, and thereby obtain figures of merit of over 45 Gc/s. Experiments are now in progress to evaluate these diodes in a Q-band amplifier.

9. Future Trends

The existing varactor diode encapsulations will probably be improved by the more extensive use of copper to increase thermal dissipation, and by the replacement of high-alumina ceramics with lower dielectric constant insulators to reduce the stray capacitance. This should allow fully encapsulated diodes to operate well above Q-band and the technique of mounting the junction element directly in the waveguide will only be required at much higher frequencies.

The main limitations on junction performance are now due to the components of the series resistance arising from the bulk resistance and skin effects. There is no doubt that the further refinement of epitaxial techniques will continue to reduce the bulk resistance and allow higher breakdown voltages for harmonic generator applications. In particular, recent experiments with the basic structures described in this paper have shown that it is possible to apply oxide masking and planar techniques to gallium arsenide and obtain satisfactory microwave varactors. Not only will this lead to further exploitations of the epitaxial structure, but also enable completely novel device geometries to be considered.

Apart from the diffused junction a re-examination of the characteristics of a metal-semiconductor contact against the background of current device techniques suggest that this form of diode could have useful properties. By employing n-type gallium arsenide in conjunction with an evaporated metal contact a truly abrupt junction is obtained, the characteristic of which can be controlled by the choice of metal. Some early experimental diodes of this type have been made in the C-band varactor structure and yielded cut-off frequencies of the order of 200 Gc/s.

10. Conclusions

It has been demonstrated that gallium arsenide is a suitable semiconductor material for the fabrication of microwave varactor diodes intended for low noise parametric amplification and harmonic generation. The implications of the evaluation of the characteristics of varactor diodes at microwave frequencies have been applied to the development of three forms of diode for use at C, X and Q-band (Fig. 14) having typical cut-off frequencies of 150, 250 and 700 Gc/s (at zero bias), respectively. With further progress in

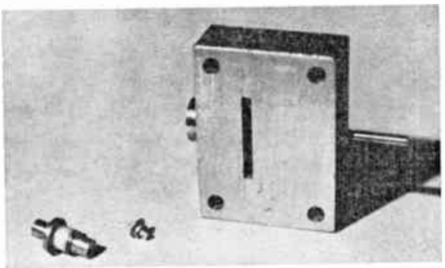


Fig. 14. From left to right: C-, X- and Q-band varactor diodes.

gallium arsenide technology, higher mobility bulk and epitaxial material will become available and by combining these with new device structures based on planar processes it should be feasible to achieve cut-off frequencies of the order of 1000 Gc/s. Diodes of this type would give significant increases in harmonic generator efficiency and would enable low noise parametric amplifiers rivalling the maser to be realized.

11. Acknowledgments

This work was supported by the Ministry of Defence whose permission to publish this paper is acknowledged.

12. References

1. C. S. Aitchison, R. Davis and P. J. Gibson, 'A simple diode parametric amplifier design for use at S-, C- and X-band', *Proceedings of the Symposium on Microwave Applications of Semiconductors*, London, 1965. Paper No. 26.
2. P. Penfield and R. P. Rafuse, 'Varactor Applications' (M.I.T. Press, Cambridge, Mass., 1962).
3. K. N. Chang, 'Parametric and Tunnel Diodes' (Prentice-Hall, Englewood Cliffs, 1964).
4. C. M. Allen, P. R. Liegey and B. Salzburg, '10°K noise temperature from indium antimonide varactors at 77°K', *Proc. Inst. Elect. Electronics Engrs*, **51**, p. 856, May 1963.
5. J. J. Fink and R. L. Rulison, 'Epitaxial silicon varactors at low temperatures and microwave frequencies', *Proc. I.E.E.E.*, **52**, p. 420, April 1964.
6. L. H. Gibbons, M. F. Lamorte and E. E. Widmer, 'High cut-off frequency GaAs diffused junction varactor diodes', *R.C.A. Review*, **24**, p. 199, June 1963.
7. D. A. E. Roberts and K. Wilson, 'Evaluation of high quality varactor diodes', *Proceedings of the Symposium on Microwave Applications of Semiconductors*, London, 1965. Paper No. 22. To be reprinted in *The Radio and Electronic Engineers*, **31**, 1966.
8. E. Denison, Admiralty Surface Weapons Establishment, Private Communication.
9. J. J. Chang, J. H. Forster and R. M. Ryder, 'Semiconductor junction varactors with high voltage sensitivity', *Trans. I.E.E.E., on Electron Devices*, **ED-10**, p. 281, July 1963.
10. C. S. Aitchison, Mullard Research Laboratories, Private Communication.
11. J. Sadler and W. Wells, 'Cryogenically cooled X-band parametric amplifier', *Proc. I.E.E.E.*, **52**, p. 1268, March 1964.
12. J. Munro and E. Feldman, 'K.A. band klystron replacement', *Trans. I.E.E.E., on Microwave Theory and Technique*, **MTT-12**, No. 5, p. 553, September 1964 (Letter).

Manuscript received by the Institution on 11th May 1965. (Paper No. 1039.)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full biographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

CARRIER REINSERTION IN TELEVISION

The problem of the right magnitude and phase of the auxiliary carrier in a carrier re-insertion method is dealt with in a German paper, assuming that the auxiliary carrier has no amplitude or phase distortion. The amplitude is determined by the desired reduction in the modulation index and the phase must be identical to the phase of the signal-carrier. No use can be made of the possibility of deviating from this phase relationship for the purpose of compensating distortions caused by the group delay characteristic because compensation can be obtained only within a narrow band of modulation frequencies and a phase shift would produce errors at other frequencies. The signal/noise ratio in the video signal becomes worse when the auxiliary carrier contains interference amplitude or phase modulation. The video signal/noise ratio is quoted as a function of the interference amplitude modulation index and as a function of the interference phase deviation of the auxiliary carrier.

Three methods are investigated for solving the problem of obtaining the auxiliary carrier from the modulated signal: limiting, selection and synchronization of an oscillator. A series of oscillograms illustrates the effect of demodulating with carrier re-insertion.

'Television signal demodulation by carrier re-insertion', H. C. Höring, *Nachrichtentechnische Zeitschrift*, **18**, No. 10, pp. 579–92, October 1965.

PROPAGATION OF M.F. SIGNALS

An Australian paper describes a method of transmission called 'orthogonal transmission' whereby medium frequency sky-wave signals are subjected to considerably greater absorption than that imposed on a vertically polarized transmission.

Under the influence of the Earth's magnetic field, a radio wave incident upon the ionosphere splits into two waves referred to as the ordinary wave and the extraordinary wave. The ordinary wave critical frequency of the E layer is the same as that calculated for no magnetic field, but that for the extraordinary wave is lower when the gyrofrequency exceeds the transmission frequency, and consequently this wave penetrates deeper into the ionosphere than the ordinary wave. Under certain conditions it is subjected to considerably greater absorption than the ordinary wave.

The suggested method of transmission is the one which changes the polarization of a medium frequency transmission from vertical to the particular elliptical form required for propagation by the extraordinary mode in the ionosphere. This ensures virtually no propagation via the ordinary mode and consequently the sky-wave signal received at the surface of the Earth would be severely attenuated due to high absorption of the extraordinary wave. Successful application of this system requires a knowledge of the conditions under which the extraordinary wave is more severely attenuated than the ordinary wave.

'The absorption of medium frequency sky-waves by close coupling to the extraordinary mode', J. M. Dixon, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **26**, No. 12, pp. 369–80, December 1965.

ANTENNA MEASUREMENTS USING RADIO ASTRONOMY TECHNIQUES

The use of contemporary large antennas has made evident the fact that the existing 'ground-based' methods for determination of parameters of these antennas have become unacceptable in a number of practical cases. Moreover, in conjunction with the development of low noise receiving systems and with improvements in the accuracy of measurements, it becomes necessary to determine accurately the efficiency and the scattering coefficient of an antenna, the measurement of which by ordinary ground-based methods is difficult and in some cases impossible. Thus the development of antenna techniques for radio astronomy has necessitated development for methods of antenna investigation which opened up new possibilities for calibration of antenna systems, which are applicable in physics and engineering investigations.

By radio astronomy methods are meant not only the methods which use extraterrestrial sources of radiation but also the methods which originated and were developed in the process of development of radio astronomy, and which are based upon the use of radio astronomy equipment. The latter include the methods which utilize the thermal radiation of an 'absolutely black body' as a standard.

The main advantage of radio astronomy methods is their relative simplicity and the fact that the extraterrestrial sources are always located in the far field of any antenna, no matter how large, and can be observed at various angles which may be very significant. In particular, these methods can be employed in the process of design and exploitation of large antennas for which the use of transmitters, elevated above ground, presents considerable difficulties. A Russian survey has analysed the problem from the point of view of applicability of developed contemporary methods of radio astronomy to investigation of antennas.

'Methods of radio astronomy in investigation of antennas (A survey)', N. M. Tseytlin, *Radio Engineering and Electronic Physics*, (English language edition of *Radioteknika i Electronika* (U.S.S.R.)), **10**, No. 8, pp. 1175–1208, August 1965.